1 GSW… Matrices

"Rows first, columns second. Remember that. R then C."[1]

A matrix is a set of real or complex numbers arranged in a rectangular array. They can be any size and shape (provided they are rectangular). A few examples:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 1 \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} 2 & 1 & -3 \\ 4 & -2 & 3 \end{bmatrix} \qquad \mathbf{E} = \begin{bmatrix} 1 & 2 \end{bmatrix} \qquad \mathbf{F} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \qquad \mathbf{H} = [2] \qquad \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The size of a matrix is usually written in terms of the number of rows and the number of columns in the rectangular array, for example matrix **C** above has three rows and two columns. It's always expressed in that order: rows first, columns second. If you're going to be using MATLAB[2], then it's a very good idea to know that.

The matrices **A**, **B**, **H** and **I** above have the same number of rows as columns, and are therefore known as *square matrices*. A *fat matrix* is one with more columns than rows, for example matrix **D** above. A *thin or skinny matrix* is one with more rows than columns, for example matrix **C**. Matrix **E** has only one row, and can also be called a *row vector*. Matrices **F** and **G** only have one column, and can be called *column vectors*. Matrix **H** just has one element, so this is just a scalar.

1.1 Notation

As you'll have noticed, I'm writing a matrix as an upper-case bold non-italic letter, like this: **A**. If I want to refer to just one element in a matrix, for example the second element in the first row, I'd write this as **A**[1,2] or $A_{1,2}$. Note row first, column second. With the matrix **A** above, $A_{1,2} = 2$. This is the conventional way to write matrices, although when doing matrix algebra and calculus, I often find a different notation more useful.

The alternate notation is to write a matrix in the form A_{ij}. If you use this notation, it's important to remember that A_{ij} is not just the term in the i^{th} row and j^{th} column, it is the whole matrix. The idea is that any index (in this case i and j) takes all of its range of values. If I wanted to refer to just one element of this matrix, I'd write it as, for example, $(A_{ij})_{1,2}$. That

[1] This isn't a famous quote, but it's so useful to know that I thought I'd write it up here in the hope that this might make it more memorable.

[2] MATLAB is a very powerful and commonly-used mathematics and simulation program that is especially suited to working with matrices. The format of the MATLAB instructions for working with matrices takes a bit of getting used to, so I'll add in some hints in the footnotes.

© 2012 Dave Pearce

might seem a bit confusing at the moment, but it can help when you try and multiply matrices together. Which brings me to the subject of matrix arithmetic.

1.2 Matrix Arithmetic

Matrices can be added, subtracted and multiplied and a form of division can be used[3], however it's not always possible to do any of these operations on two matrices. Which operations are possible with any two given matrices depends on the size of the matrices.

1.2.1 Matrix Addition and Subtraction

Adding and subtracting matrices can only be done when the two matrices are the same size, and then the sum of two matrices is a matrix with each element the sum of the corresponding terms in the component matrices. For example:

$$\mathbf{A}+\mathbf{B}=\mathbf{C} \qquad A_{ij}+B_{ij}=C_{ij}$$

$$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}+\begin{bmatrix}5 & 6\\7 & 8\end{bmatrix}=\begin{bmatrix}1+5 & 2+6\\3+7 & 4+8\end{bmatrix}=\begin{bmatrix}6 & 8\\10 & 12\end{bmatrix} \tag{0.1}$$

Subtraction works the same way:

$$\mathbf{A}-\mathbf{B}=\mathbf{C} \qquad A_{ij}-B_{ij}=C_{ij}$$

$$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}-\begin{bmatrix}5 & 6\\7 & 8\end{bmatrix}=\begin{bmatrix}1-5 & 2-6\\3-7 & 4-8\end{bmatrix}=\begin{bmatrix}-4 & -4\\-4 & -4\end{bmatrix} \tag{0.2}$$

Addition and subtraction are being done on a *term-by-term* basis, so for each term in **A**, there must be a corresponding term both in **B**, and in the answer **C**. Since each term is being added and subtracted just like a scalar quantity, matrix addition and subtraction share the same properties as normal (scalar) addition and subtraction, for example:

$$\begin{gathered}\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}\\ (\mathbf{A}+\mathbf{B})-\mathbf{C}=\mathbf{A}+(\mathbf{B}-\mathbf{C})\\ \mathbf{A}-\mathbf{B}=-(\mathbf{B}-\mathbf{A})\end{gathered} \tag{0.3}$$

[3] I once had a student who got very cross when I suggested that it might be possible to divide matrices. If you're one of these people, try entering matrices **A** and **B** above in MATLAB, and then asking it what **A** / **B** is. You'll find it's quite happy to tell you that:

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

What it's actually doing is taking the inverse of matrix B, and then multiplying matrix A by the inverse of matrix B. It's not standard mathematical notation, but it's unambiguous and clearly defined, so I don't see any problem with it myself.

DRAFT

1.2.2 Matrix Multiplication

Addition and subtraction are possible provided the two matrices involved are exactly the same shape and size. Multiplication, on the other hand, is only possible if the *inner dimension* of the two matrices is the same[4].

The term 'inner dimension' makes more sense if you write the matrices in the alternate form, as, for example, A_{ij} and B_{kl}. The 'inner dimensions' are the dimensions next to each other (on the 'inside'), in this case the j of A_{ij} and the k of B_{kl}. (Note j is an index that tells you what column the element A_{ij} is in, k is an index that tells you what row the element of B_{kl} is on.) If these dimensions are equal (in other words if the number of columns in **A** is equal to the number of rows in **B**), then the two matrices can be multiplied together, and the result is a matrix with a number of rows equal to the number of rows in **A**, and a number of columns equal to the number of columns in **B**. For example, a 3-by-2 matrix multiplied by a 2-by-5 matrix, results in a 3-by-5 matrix. That's three rows, and five columns.

The terms in the product matrix are the sum of the products of the terms in the corresponding row of the first matrix and column of the second matrix. For example:

$$\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix}\begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix}=\begin{bmatrix} a_{1,1}b_{1,1}+a_{1,2}b_{2,1} & a_{1,1}b_{1,2}+a_{1,2}b_{2,2} \\ a_{2,1}b_{1,1}+a_{2,2}b_{2,1} & a_{2,1}b_{1,2}+a_{2,2}b_{2,2} \end{bmatrix} \tag{0.4}$$

We could write the term $C_{i,j}$ in the product matrix as:

$$C_{i,j}=\sum_{k=1}^{N} A_{i,k}\,B_{k,j} \tag{0.5}$$

where N is both the number of columns in **A** and the number of rows in **B** (which must be the same, otherwise multiplication doesn't work). For those familiar with the concept of a vector dot product, it might help to think of the matrix **A** as composed of row vectors, and the matrix **B** to be composed of column vectors. Then the product **AB** is a matrix **C** with terms $C_{i,j}$ which are the dot (inner) product of the row vector in the i^{th} row of **A** and the column vector in the j^{th} column of **B**.

For example, a 3-by-2 matrix (3 rows, 2 columns) multiplied by a 2-by-3 matrix (2 rows, three columns) gives a 3-by-3 matrix:

$$\mathbf{CD}=\begin{bmatrix} 1 & 3 \\ 3 & 4 \\ 5 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 & -3 \\ 4 & -2 & 3 \end{bmatrix}=\begin{bmatrix} 2+12 & 1-6 & -3+9 \\ 6+16 & 3-8 & -9+12 \\ 10+4 & 5-2 & -15+3 \end{bmatrix}=\begin{bmatrix} 14 & -5 & 6 \\ 22 & -5 & 3 \\ 14 & 3 & -12 \end{bmatrix} \tag{0.6}$$

whereas the same two matrices multiplied together in the reverse order gives:

[4] There's an exception to this rule. If one of the matrices is a 1-by-1 scalar, then this scalar can multiply any matrix of any size and shape. The result is a matrix of the same size and shape, with every element multiplied by the scalar. Division by a scalar works in the same term-by-term way.

$$\mathbf{DC} = \begin{bmatrix} 2 & 1 & -3 \\ 4 & -2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 3 & 4 \\ 5 & 1 \end{bmatrix} = \begin{bmatrix} 2+3-15 & 6+4-3 \\ 4-6+15 & 12-8+3 \end{bmatrix} = \begin{bmatrix} -10 & 7 \\ 13 & 7 \end{bmatrix} \tag{0.7}$$

Because the order is important, it must be clearly specified. The product **CD** is known as **D** *pre-multiplied* by **C**, or **C** *post-multiplied* by **D**.

Vectors are a special form of matrix with either one column (column vectors) or one row (row vectors). A couple of vector examples:

A 3-by-1 column vector multiplied by a 1-by-2 row vector gives a 3-by-2 matrix:

$$\mathbf{GE} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ -1 & -2 \end{bmatrix} \tag{0.8}$$

but a 1-by-2 row vector multiplied by a 2-by-1 column vector gives a 1-by-1 scalar:

$$\mathbf{EF} = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 3+2 \end{bmatrix} = \begin{bmatrix} 5 \end{bmatrix} \tag{0.9}$$

Vectors are often just referred to only in terms of the number of elements they contain, and this makes it impossible to tell whether a vector is a row vector or a column vector, and therefore what happens when two of them are multiplied together using matrix multiplication. In this book, all vectors are column vectors unless otherwise stated. If I want a row vector, I'll define it as the *transpose*[5] of a column vector, for example:

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}^T \tag{0.10}$$

1.2.3 The Alternate Notation[6]

Using the alternate notation I mentioned, a matrix **A** is written as A_{ij}. Multiplying two matrices together then produces a result:

$$C_{ik} = A_{ij}\, B_{jk} \tag{0.11}$$

where the fact that the index j appears in both A_{ij} and B_{jk} implies that there is a summation being done over this index. In full, this could be written:

[5] See later in this chapter for more about the transpose of a matrix.

[6] Note: it's safe to ignore this section if it seems confusing. You only need to understand one notation, and the first one introduced in section 1.2.2 is the more common.

$$C_{ik} = \sum_{j=1}^{N} A_{ij}\, B_{jk} \tag{0.12}$$

although we don't usually write the summation sign, we just assume that if any index appears more than once in any product term, it must be summed over. Using this notation, the order of writing the matrices doesn't matter, what's important is the order of the indices. To work out what a product implies in the more conventional notation, all you need to do is arrange the terms so that the shared index is the 'inner dimension' first. For example:

$$D_{jk} C_{ij} = C_{ij} D_{jk} = \mathbf{CD} \tag{0.13}$$

This also removes the problem with multiplying vectors. If we have two vectors **E** and **F**, then $E_i F_i$ is the dot-product of the two vectors:

$$E_i F_i \Rightarrow \sum_{i=1}^{N} E_i F_i \tag{0.14}$$

whereas $E_i F_j$ is a matrix[7].

Multiplying a matrix with a vector can equally provide one of two results, depending on the order of the multiplication:

$$\begin{gathered} \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -2 & -1 \end{bmatrix} = \begin{bmatrix} 1-4 & 3-2 \end{bmatrix} = \begin{bmatrix} -3 & 1 \end{bmatrix} \\ \begin{bmatrix} 1 & 3 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1+6 \\ -2-2 \end{bmatrix} = \begin{bmatrix} 7 \\ -4 \end{bmatrix} \end{gathered} \tag{0.15}$$

In terms of this notation, $x_j A_{jk} = A_{jk} x_j$ is a row vector post-multiplied by a matrix ($\mathbf{x}^T\mathbf{A}$ in conventional notation), and results in a row vector, $A_{jk} x_k = x_k A_{jk}$ is a column vector pre-multiplied by a matrix (in other terms $\mathbf{Ax}$), and results in another column vector.

1.2.4 Unit Matrices

It is sometimes useful to have a matrix that does not change the value of any matrix that is multiplied by it, just like any scalar number multiplied by one doesn't change its value. This so-called *unit matrix* is always square, and has all elements on its leading diagonal (the one from top-left to bottom-right) equal to one, and all other elements equal to zero. For example, the 3-by-3 unit matrix looks like this:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

It's straightforward to prove that this doesn't change the value of any matrix that is either pre-multiplied or post-multiplied by it: consider any element $\mathbf{B}_{i,k}$ in a matrix **B** multiplied by the unit matrix:

[7] Known as the *outer product* of the two vectors.

$$\sum_j \mathbf{I}_{i,j}\mathbf{B}_{j,k} = \mathbf{I}_{i,i}\mathbf{B}_{i,k} = \mathbf{B}_{i,k} \tag{0.16}$$

(note that in the summation, the only value of **I** that is non-zero is the one which both indices are the same, there is only one of these each summation, and the value of $I_{i,i}$ here is one).

A few examples:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 3+0+0 \\ 0+2+0 \\ 0+0+5 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix} \tag{0.17}$$

$$\begin{aligned} \begin{bmatrix} 3 & 2 & 5 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} &= \begin{bmatrix} 3+0+0 & 0+2+0 & 0+0+5 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 2 & 5 \end{bmatrix} \end{aligned} \tag{0.18}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 & -2 \\ -1 & -3 & 4 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 & -2 \\ -1 & -3 & 4 \\ 0 & 1 & 0 \end{bmatrix} \tag{0.19}$$

1.2.5 Inverse Matrices and Division

It is usually (although not always) possible to produce an inverse matrix $\mathbf{A}^{-1}$, such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$, where **I** is the unit matrix[8], provided the matrix **A** is square. For example, the inverse matrix of **A** above is:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^{-1} = \frac{1}{2}\begin{bmatrix} -4 & 2 \\ 3 & -1 \end{bmatrix} \tag{0.20}$$

since,

[8] In terms of the subscript notation, **I** can be written as δ_{ij}, where the special matrix δ_{ij} has a value of one when its indices are equal and a value of zero when they are not.

$$\begin{aligned}\mathbf{AA}^{-1} &= \begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}\frac{1}{2}\begin{bmatrix}-4 & 2\\3 & -1\end{bmatrix}\\ &= \frac{1}{2}\begin{bmatrix}1\times(-4)+2\times3 & 1\times2+2\times(-1)\\3\times(-4)+4\times3 & 3\times2+4\times(-1)\end{bmatrix}\\ &= \frac{1}{2}\begin{bmatrix}-4+6 & 2-2\\-12+12 & 6-4\end{bmatrix}\\ &= \frac{1}{2}\begin{bmatrix}2 & 0\\0 & 2\end{bmatrix} = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\end{aligned} \tag{0.21}$$

In general, working out the inverse of a matrix is a time-consuming process (although it's a common requirement). In the particular case of a 2-by-2 matrix there is a simple formula for the inverse:

$$\begin{bmatrix}a & b\\c & d\end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix}d & -b\\-c & a\end{bmatrix} \tag{0.22}$$

Note that if $ad = bc$ then the inverse doesn't exist, any attempt to calculate it would result in trying to divide by zero. A square matrix that does have an inverse is known as an *invertible matrix*. A square matrix that does not have an inverse is known as *singular* or *degenerate*. Finding efficient techniques for working out the inverse of large matrices (and avoiding having to work them out) is an active and fruitful area for communications engineering, and an introduction to some of these techniques are subject of much of the remainder of this section of the book.

'Division' can then be defined in terms of multiplying by a matrix inverse. For example, if:

$$\mathbf{AB} = \mathbf{C} \tag{0.23}$$

and **B** and **C** are known and **B** is square and invertible, and you want to find the matrix **A**, then:

$$\begin{aligned}\mathbf{ABB}^{-1} &= \mathbf{CB}^{-1}\\ \mathbf{A} &= \mathbf{CB}^{-1}\end{aligned} \tag{0.24}$$

since $\mathbf{B}^{-1}\mathbf{B} = \mathbf{I}$ and $\mathbf{AI} = \mathbf{A}$.[9]

1.2.6 Complex Conjugates

This one's very easy. The complex conjugate of a matrix is another matrix of the same shape and size with each term the complex conjugate of the original. For example:

$$\begin{bmatrix}1+j & 2 & -3j\\0 & 1-j & 1\end{bmatrix}^{*} = \begin{bmatrix}1-j & 2 & 3j\\0 & 1+j & 1\end{bmatrix} \tag{0.25}$$

[9] Calculating $\mathbf{CB}^{-1}$ is what MATLAB will do if you ask it to divide two matrices using an expression of the form `C/B`. Also, in MATLAB, `A\C` means $\mathbf{A}^{-1}\mathbf{C}$. It's a common mistake to get the "/" and "\" mixed up.

I'll write the complex conjugate of a matrix **A** as $\mathbf{A}^*$, just like for scalars. (Warning: this is not a universally accepted notation.)

1.2.7 Transposition

Apart from addition, subtraction, multiplication and 'division', there is one more common matrix operation, and this one does not have a parallel in the more familiar world of scalars. Transposition, written as $\mathbf{A}^T$, involves taking the matrix **A** and swapping over the rows and columns, so that the rows in **A** become the columns in $\mathbf{A}^T$ and vice versa.

If **A** has r rows and c columns, then $\mathbf{A}^T$ has c rows and r columns. For example:

$$\mathbf{A}_{i,j}{}^T = \mathbf{A}_{j,i} \tag{0.26}$$

$$\begin{bmatrix} 1 & 2 & 4 \\ 3 & 5 & 7 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 1 & 3 \\ 2 & 5 \\ 4 & 7 \end{bmatrix} \tag{0.27}$$

A symmetric matrix is a matrix that is equal to its own transpose, so that that $A_{ij} = A_{ji}$. Obviously (at least I hope it's obvious), all symmetric matrices are square.

There's also a very important variation on transposition, which involves taking the complex conjugate of the transpose of a matrix. This turns out to be so useful that it has its own name: the *Hermitian transpose*[10] (also known as the *conjugate transpose*). The Hermitian transpose of a matrix **A**, written as $\mathbf{A}^H$, is the complex conjugate of the transpose[11]. For example:

$$\mathbf{A}_{i,j}{}^H = \mathbf{A}_{j,i}{}^* \tag{0.28}$$

$$\begin{bmatrix} 1+j & 2 & 4 \\ 3 & 5-j & 7+2j \end{bmatrix}^H = \begin{bmatrix} 1-j & 3 \\ 2 & 5+j \\ 4 & 7-2j \end{bmatrix} \tag{0.29}$$

A *Hermitian matrix* is one that is equal to its own Hermitian transpose. Obviously (again I hope it's obvious), all Hermitian matrices are square; and slightly less obviously, all the terms on the main diagonal of a Hermitian matrix are real.

1.3 Matrix Algebra

Addition and subtraction with matrices works just like with scalars, but multiplication is different, and using the new operation of transposition requires some care when solving algebraic expressions including matrices. Firstly, it's useful to know a few useful results about the transposes, inverses and Hermitian transposes of products of matrices.

[10] Named after a Charles Hermite, a 19th century French mathematician.

[11] Some texts use the notation $\mathbf{A}^*$ to indicate the Hermitian or conjugate transpose of **A**. I don't like this notation, since it can be easily confused with the complex conjugate of **A**. Note that in MATLAB `A'` is the Hermitian transpose of **A**, not the transpose. If you want a transpose in MATLAB, you have to ask for `A.'` (note the additional decimal point). If you want the complex conjugate, you have to ask for `conj(A)`, or, if you're feeling clever, `(A.')'`.

© 2012 Dave Pearce

Consider the transpose of a matrix product $(\mathbf{AB})^T$. It can be easily shown that this is equal to $\mathbf{B}^T\mathbf{A}^T$ by considering the value of each element of $\mathbf{AB}$.

On the j[th] row and i[th] column of $\mathbf{B}^T\mathbf{A}^T$ is the result of adding up the product of the terms on the j[th] row of $\mathbf{B}^T$ and the i[th] column of $\mathbf{A}^T$. But the terms on the j[th] row of $\mathbf{B}^T$ are the terms on the j[th] column of $\mathbf{B}$, and the terms on the i[th] column of $\mathbf{A}^T$ are the terms on the i[th] row of $\mathbf{A}$.

The value of the i[th] row and j[th] column of $\mathbf{AB}$ is also the sum of the product of the terms on the i[th] row of $\mathbf{A}$ and the j[th] column of $\mathbf{B}$. So the value on the j[th] row and i[th] column of $\mathbf{B}^T\mathbf{A}^T$ is the same as the value on the i[th] row and j[th] column of $\mathbf{AB}$. In other words: $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.

Alternatively, using the suffix notation, for any matrix $\mathbf{C}$, the value of $C_{i,k}$ must be equal to the value of $C^T_{k,i}$, this follows directly from the definition of the transpose. Therefore: $(\mathbf{AB})^T_{i,k} = (\mathbf{AB})_{k,i}$. Taking the transpose is equivalent to just swopping over the order of the indices.

Expressing this term in the multiplication as a summation of the relevant terms gives:

$$(\mathbf{AB})^T{}_{i,k} = \left(\sum_j \mathbf{A}_{i,j}\mathbf{B}_{j,k}\right)^T = \left(\sum_j \mathbf{A}_{k,j}\mathbf{B}_{j,i}\right) = \left(\sum_j \mathbf{B}_{i,j}{}^T\mathbf{A}_{j,k}{}^T\right) = \left(\mathbf{B}^T\mathbf{A}^T\right)_{i,k} \quad (0.30)$$

For example:

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

and

$$(\mathbf{BA})^T = \left(\begin{bmatrix} 1 & -2 & 0 \\ 2 & 1 & -4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -1 & 0 \\ 0 & 1 \end{bmatrix}\right)^T = \begin{bmatrix} 3 & 2 \\ 1 & 0 \end{bmatrix}^T = \begin{bmatrix} 3 & 1 \\ 2 & 0 \end{bmatrix}$$

Likewise, the Hermitian transpose of a matrix product $(\mathbf{AB})^H$ is equal to $\mathbf{B}^H\mathbf{A}^H$.

$$(\mathbf{AB})^H = \mathbf{B}^H\mathbf{A}^H \quad (0.31)$$

and a similar result sometimes applies to the inverse of a product of matrices:

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \quad (0.32)$$

since:

$$(\mathbf{AB})(\mathbf{AB})^{-1} = \mathbf{I} = \mathbf{AA}^{-1} = \mathbf{A}\left(\mathbf{BB}^{-1}\right)\mathbf{A}^{-1} = (\mathbf{AB})\left(\mathbf{B}^{-1}\mathbf{A}^{-1}\right) \quad (0.33)$$

Note the word 'sometimes'. This doesn't always work. Just because the product $\mathbf{AB}$ is invertible doesn't necessarily mean that $\mathbf{A}^{-1}$ or $\mathbf{B}^{-1}$ even exist: for example, if $\mathbf{A}$ is a 2-by-3 matrix, and $\mathbf{B}$ is a 3-by-2 matrix, then the product $\mathbf{AB}$ is a square 2-by-2 matrix and could well be invertible, but since $\mathbf{A}$ and $\mathbf{B}$ are not square, they don't have an inverse.

DRAFT

Given a formula like:

$$\mathbf{y} = \mathbf{A}\left(\mathbf{x} + \mathbf{B}\mathbf{x}\right) \tag{0.34}$$

where **x** and **y** are column vectors, and **A** and **B** are square matrices with the same number of columns as **x** has rows, we often need to solve for **x**. Using the concept of the matrix inverse, this is not too hard:

$$\begin{aligned} \mathbf{y} &= \mathbf{A}\left(\mathbf{x} + \mathbf{B}\mathbf{x}\right) \\ \mathbf{A}^{-1}\mathbf{y} &= \mathbf{A}^{-1}\mathbf{A}\left(\mathbf{x} + \mathbf{B}\mathbf{x}\right) \\ \mathbf{A}^{-1}\mathbf{y} &= \mathbf{x} + \mathbf{B}\mathbf{x} \\ &= \left(\mathbf{I} + \mathbf{B}\right)\mathbf{x} \\ \left(\mathbf{I} + \mathbf{B}\right)^{-1}\mathbf{A}^{-1}\mathbf{y} &= \left(\mathbf{I} + \mathbf{B}\right)^{-1}\left(\mathbf{I} + \mathbf{B}\right)\mathbf{x} \\ \mathbf{x} &= \left(\mathbf{I} + \mathbf{B}\right)^{-1}\mathbf{A}^{-1}\mathbf{y} \end{aligned} \tag{0.35}$$

Note that I replaced **x** + **Bx** with (**I** + **B**)**x** using the properties of the unit matrix: anything multiplied by the unit matrix gives itself. The unit matrix is always square, but it can be whatever size is required to allow the multiplication to be done. A quick reminder example:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} a+0+0 \\ 0+b+0 \\ 0+0+c \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \tag{0.36}$$

here the unit matrix is multiplying a column vector with three component, and is therefore a square with three rows and three columns. If it was multiplying a row vector with three components, the unit matrix would be a scalar (a one-by-one matrix): in other words just the number one.

1.3.1 Some Matrix Identities and Theorems

There are a large number of matrix identities, most of which are straightforward to prove (see the problems for a few examples). One that isn't so easy to prove, but turns out to be very useful, is known as the Matrix Inversion Lemma. This states that:

$$\left(\mathbf{A}^{-1} + \mathbf{U}\mathbf{V}^{H}\right)^{-1} = \mathbf{A} - \mathbf{A}\mathbf{U}\left(\mathbf{I} + \mathbf{V}^{H}\mathbf{A}\mathbf{U}\right)^{-1}\mathbf{V}^{H}\mathbf{A} \tag{0.37}$$

This turns out to be a very useful formula, since it shows what happens to a matrix when a small amount (in this case $\mathbf{U}\mathbf{V}^{H}$) is added to its inverse[12]. The easiest way to show this formula is true is to multiply the matrix ($\mathbf{A}^{-1}$ + $\mathbf{U}\mathbf{V}^{H}$) with the right-hand-side of equation (0.37) and check that the result is the unit matrix.

[12] You might be wondering why this formula is expressed in terms of $\mathbf{V}^{H}$, and not just **V**. It's to do with the fact that in some applications of this formula, **U** and **V** are not square matrices, but vectors. Consider the left-hand side of the equation: to make a matrix that you can add to a square matrix (remember that all inverse matrices are square), you need to create the outer product of the two vectors, and that means post-multiplying a column vector by a row vector. Since all vectors are by default column vectors, we create $\mathbf{V}^{H}$, which being the Hermitian transpose of a column vector is a row vector (with terms equal to the complex conjugates of the terms in the column vector **V**).

If ($\mathbf{A}^{-1} + \mathbf{U}\mathbf{V}^H$) is square (which it must be, since $\mathbf{A}^{-1}$ is an inverse matrix, and therefore must be square itself, and you can only add matrices that are the same size and shape) then anything that multiplies ($\mathbf{A}^{-1} + \mathbf{U}\mathbf{V}^H$) to give the unit matrix must be the inverse of ($\mathbf{A}^{-1} + \mathbf{U}\mathbf{V}^H$).

Note that here:

$$\begin{aligned}
&\left(\mathbf{A}^{-1}+\mathbf{U}\mathbf{V}^H\right)\left(\mathbf{A}-\mathbf{A}\mathbf{U}\left(\mathbf{I}+\mathbf{V}^H\mathbf{A}\mathbf{U}\right)^{-1}\mathbf{V}^H\mathbf{A}\right)\\
&=\mathbf{A}^{-1}\mathbf{A}+\mathbf{U}\mathbf{V}^H\mathbf{A}-\mathbf{A}^{-1}\mathbf{A}\mathbf{U}\left(\mathbf{I}+\mathbf{V}^H\mathbf{A}\mathbf{U}\right)^{-1}\mathbf{V}^H\mathbf{A}-\mathbf{U}\mathbf{V}^H\mathbf{A}\mathbf{U}\left(\mathbf{I}+\mathbf{V}^H\mathbf{A}\mathbf{U}\right)^{-1}\mathbf{V}^H\mathbf{A}\\
&=\mathbf{I}+\mathbf{U}\mathbf{V}^H\mathbf{A}-\mathbf{U}\left(\mathbf{I}+\mathbf{V}^H\mathbf{A}\mathbf{U}\right)^{-1}\mathbf{V}^H\mathbf{A}-\mathbf{U}\mathbf{V}^H\mathbf{A}\mathbf{U}\left(\mathbf{I}+\mathbf{V}^H\mathbf{A}\mathbf{U}\right)^{-1}\mathbf{V}^H\mathbf{A}\\
&=\mathbf{I}+\mathbf{U}\mathbf{V}^H\mathbf{A}-\mathbf{U}\left(\mathbf{I}+\mathbf{V}^H\mathbf{A}\mathbf{U}\right)\left(\mathbf{I}+\mathbf{V}^H\mathbf{A}\mathbf{U}\right)^{-1}\mathbf{V}^H\mathbf{A}\\
&=\mathbf{I}+\mathbf{U}\mathbf{V}^H\mathbf{A}-\mathbf{U}\mathbf{V}^H\mathbf{A}=\mathbf{I}
\end{aligned}\tag{0.38}$$

Since multiplying ($\mathbf{A}^{-1} + \mathbf{U}\mathbf{V}^H$) with the right-hand-side of equation (0.37) gives the unit matrix, this must have been the inverse of the matrix.

1.3.2 Some Common Mistakes

Matrices can be added, subtracted and multiplied just like ‘normal’ scalar quantities, but there are some situations in which they don’t behave in quite the same way. A few to watch out for:

- One we’ve seen already: $(\mathbf{AB})^{-1}$ is not necessarily equal to $\mathbf{B}^{-1}\mathbf{A}^{-1}$. Just because **AB** is invertible, doesn’t mean that either **A** or **B** is invertible, or even square. For example:

$$\begin{bmatrix}1 & 2 & -1\\ 2 & -1 & 2\end{bmatrix}\begin{bmatrix}1 & -1\\ 2 & 1\\ -3 & 1\end{bmatrix}=\begin{bmatrix}8 & 0\\ -6 & -1\end{bmatrix}\tag{0.39}$$

  and,

$$\begin{bmatrix}8 & 0\\ -6 & -1\end{bmatrix}^{-1}=\frac{1}{8}\begin{bmatrix}1 & 0\\ -6 & -8\end{bmatrix}\tag{0.40}$$

  but neither of the matrices **A** or **B** can have an inverse since they’re not square.

- If you have an equation **AB** = 0, then you might think that either **A** or **B** must be a matrix of zeros. If so, you’d be wrong. For example:

$$\begin{bmatrix}0 & 2 & 2 & 1\\ 2 & 0 & 2 & 4\end{bmatrix}\begin{bmatrix}1 & 2\\ 1 & -1\\ -1 & 2\\ 0 & -2\end{bmatrix}=\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}\tag{0.41}$$

What you can say is that if **A** and **B** are square, then either **A** or **B** (or both) must have a determinant[13] of zero.

- If you have an equation of the form $\mathbf{y} = \mathbf{Ax}$, then you might think that $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$. In general, this is also wrong. It is only true if **A** is a square invertible matrix in which case vectors **x** and **y** have the same number of elements. Otherwise you can't invert the matrix, $\mathbf{A}^{-1}$ doesn't exist.

- Similarly, if you are asked to find the minimum value of $\|\mathbf{y} - \mathbf{Ax}\|^2$ where **y** and **A** are known, you might think that the answer is zero, and occurs when $\mathbf{y} = \mathbf{Ax}$. Again, in general, you'd be wrong. This is only true when the formula $\mathbf{y} = \mathbf{Ax}$ has a solution, and very often it doesn't. If the matrix **A** is invertible then there will be a solution for $\mathbf{y} = \mathbf{Ax}$, and there could be solutions even when **A** is not invertible for certain values of **y**, but in general there won't be[14].

- If you find that $\mathbf{AB} = \mathbf{AC}$, or $\mathbf{BA} = \mathbf{CA}$, then you might think that this implies that $\mathbf{B} = \mathbf{C}$. Once again, in general, you'd be wrong. For example:

$$\begin{bmatrix} 2 & -2 & -2 \\ 2 & -1 & -5 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 1 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 3 & -2 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 1 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 6 & -5 \end{bmatrix} \quad (0.42)$$

- Finally, suppose you have a matrix **B** such that $\mathbf{AB} = \mathbf{I}$. Does this mean that **B** is the inverse matrix of **A,** and that therefore $\mathbf{BA} = \mathbf{I}$? Well, not always, no. This is only true when **A** and **B** are square matrices. For example:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I} \quad (0.43)$$

but:

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & -2 \\ 0 & 1 & -1 \end{bmatrix} \quad (0.44)$$

which is obviously not a unit matrix.

1.4 Linear Simultaneous Equations

The single most useful use of matrices in communication engineering is to make writing and solving sets of linear simultaneous equations easier.

[13] See the chapter on eigenvalues and eigenvectors for more about determinants.

[14] See the chapter on linear algebra for more discussion about when this equation has a solution.

For example, consider the set of linear[15] simultaneous equations:

$$\begin{aligned} 2x+3y-z&=5\\ x-2y+z&=2\\ 3x-y+2z&=1 \end{aligned}$$

This can be written in matrix notation as:

$$\begin{bmatrix} 2 & 3 & -1 \\ 1 & -2 & 1 \\ 3 & -1 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}=\begin{bmatrix} 5 \\ 2 \\ 1 \end{bmatrix}$$

and if the 3-by-3 matrix can be inverted, then this can be solved:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}=\begin{bmatrix} 2 & 3 & -1 \\ 1 & -2 & 1 \\ 3 & -1 & 2 \end{bmatrix}^{-1}\begin{bmatrix} 5 \\ 2 \\ 1 \end{bmatrix}=\begin{bmatrix} 3 \\ -2 \\ -5 \end{bmatrix}$$

Therefore, $x = 3$, $y = -2$ and $z = -5$. All you have to do is find a fast, accurate way of inverting matrices, and you can solve any set of linear simultaneous equations. Unfortunately, finding a fast, accurate way of inverting matrices in the general case is not easy, and several numerical techniques have been developed (and new ones continue to be developed) to try and speed things up. Developing such techniques is one of the main problems of linear algebra.

1.5 Problems

1) If $\mathbf{A}=\begin{bmatrix} 3 & -2 \\ -1 & 2 \end{bmatrix}$, what is the inverse $\mathbf{A}^{-1}$? Prove that in this case $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$.

2) If,

$$\mathbf{C}=\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 1 \end{bmatrix} \qquad \mathbf{D}=\begin{bmatrix} 2 & 1 & -3 \\ 4 & -2 & 3 \end{bmatrix}$$

$$\mathbf{E}=\begin{bmatrix} 1 & 2 \end{bmatrix} \qquad \mathbf{F}=\begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Evaluate any of the following that exist:

a) $\mathbf{CD}$ b) $\mathbf{DC}$ c) $\mathbf{C}^{\mathrm{T}}+\mathbf{D}$

d) $\mathbf{F}^{\mathrm{T}}-\mathbf{E}$ e) $\mathbf{FC}$ f) $\mathbf{F}^{\mathrm{T}}\mathbf{C}$

[15] They are called linear because there no powers of the unknown variables involved. If one of the equations had a term in x^2, then these would be a set of quadratic simultaneous equations, and a lot harder to solve.

g) $\mathbf{CF}$ h) $\mathbf{C}^{T}\mathbf{F}$ i) $(\mathbf{FE})^{-1}$

j) $(\mathbf{EF})^{-1}$

3) If,

$$\mathbf{G} = \begin{bmatrix} 0 & -j \\ 1 & j \end{bmatrix} \qquad \mathbf{H} = \begin{bmatrix} 1+j & 1-j \\ 1-j & 1+j \end{bmatrix}$$

$$\mathbf{I} = \begin{bmatrix} 1 & 2j \end{bmatrix} \qquad \mathbf{J} = \begin{bmatrix} 1+3j \\ 1-j \end{bmatrix}$$

Evaluate any of the following that exist:

a) $(\mathbf{G})^{-1}$ b) $(\mathbf{H}^{2})^{-1}$ c) $\mathbf{GH}$

d) $\mathbf{GH} - \mathbf{HG}$ e) $\mathbf{GI}$ f) $\mathbf{IHJ}$

g) $\mathbf{J}^{H}\mathbf{H}$ h) $\mathbf{C}^{T}\mathbf{F}$ i) $\mathbf{J}^{T}\mathbf{G}^{-1}\mathbf{I}^{H}$

j) $\mathbf{I}^{H}\mathbf{J}^{T}$ k) $\mathbf{J}^{H}\mathbf{I}^{T}$ l) $(\mathbf{J}^{T})^{H}\mathbf{G}$

4) If you're told that $\mathbf{AB} - \mathbf{A}^{H}\mathbf{B} = 0$, what can you say about the matrix $\mathbf{A}$?

5) Prove that $(\mathbf{ABC})^{H} = \mathbf{C}^{H}\mathbf{B}^{H}\mathbf{A}^{H}$.

6) Simplify the expression $(\mathbf{AB})^{-1}\mathbf{y} - \mathbf{B}^{-1}\mathbf{x} = 0$. What about if $(\mathbf{AB})\mathbf{y} - \mathbf{Bx} = 0$, could this be simplified?

7) Simplify $\mathbf{B}(\mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{AB}^{-1})\mathbf{B}$

8) Simplify $\mathbf{ABA}^{-1} + \mathbf{A}(\mathbf{I} - \mathbf{B})\mathbf{A}^{-1}$

9) Simplify $\mathbf{B} - \mathbf{B}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B}$, assuming $\mathbf{A}$ and $\mathbf{B}$ are both invertible (note that just because ($\mathbf{A}$ + $\mathbf{B}$) is invertible, doesn't mean that $\mathbf{A}$ and $\mathbf{B}$ are individually invertible.)

10) If the inverse of $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is the matrix $\mathbf{A}^{-1} = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$, what are the inverses of:

a) $\begin{bmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ b) $\begin{bmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & 2a & 2b \\ 0 & 0 & 2c & 2d \end{bmatrix}$ c) $\begin{bmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & a & c \\ 0 & 0 & b & d \end{bmatrix}$

in terms of e, f, g and h?