Linear Algebra Reivew

One of the biggest problems people run into in this course is that quite often they have little or no background in linear algebra. Quantum mechanics is **all** linear algebra, so this is a fairly serious weakness. This review is intended to help get people who don't have a strong background get up to speed in linear algebra.

First, I'll list several websites that might be a good resource to help you assimilate this information; they have the benefit of offering a different take on the material (which is always helpful when learning something new) and also some exercises that will help you get a working knowledge of what is going on.

---

http://numericalmethods.eng.usf.edu/matrixalgebrabook/downloadma/index.html
This is a pretty good online resource for someone who either does not remember their Linear Algebra course or never took one. It goes through things step-by step and covers the basic things we are interested in. However, it is written with underclassmen in mind, so you are going to be light years ahead of where the author expects you to be in other areas. As a result, some of the introductory stuff is a little condescending. Chapters 1, 2, 3 and 10 are the relevant parts.

http://www.ling.upenn.edu/courses/Spring_2003/ling525/linear_algebra_review.html
This is an exceedingly brief review of linear algebra for people who have had a course before but need to brush up on the lingo. At the beginning, this site does a good job of going step-by-step, but later it starts to make some big leaps. The positive point is it is loosely tied to using MATLAB for illustrations, which may prove useful. You can skip all the info about SVD.

http://courses.cs.tamu.edu/rgutier/cs790_w02/l3.pdf
This site gives a short summary of the definitions of unitary matrices, normal matrices, linear independence, etc. Might be a useful reference once you've got things down.

http://matrixanalysis.com/DownloadChapters.html
This site is more advanced than the others, but bridges the gap between basic linear algebra and the types of manipulations we will go through in class. If you get even the majority of this stuff, you will be fine. Sections 3.1-3.6, 4.1-4.4, 4.7-4.8, 5.1-5.7, 7.1-7.3 and 7.5-7.6 are relevant.

So, first things first: a **Matrix** is an *n* by *m* array of numbers that we will usually display as:

$$\mathbf{M} = \overbrace{\left.\begin{pmatrix} M_{11} & M_{12} & M_{13} & \\ M_{21} & M_{22} & & \\ M_{31} & & & \\ & & & \\ & & & \end{pmatrix}\right\}}^{n} m$$

We will typically denote matrices by capital boldface letters. We refer to the matrix as having *n* **columns** and *m* **rows**. The **Transpose** of a matrix is obtained by reflecting the elements about the diagonal:

$$\mathbf{M}^T = \overbrace{\left.\begin{pmatrix} M_{11} & M_{21} & M_{31} & \\ M_{12} & M_{22} & & \\ M_{13} & & & \\ & & & \\ & & & \end{pmatrix}\right\}}^{m} n$$

Notice that the transpose is an *m* by *n* array of numbers. If one has a *p* by *q* matrix, $\mathbf{A}$, and a *q* by *r* matrix, $\mathbf{B}$, then one can multiply them:

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

which results in a *p* by *r* matrix, $\mathbf{C}$, with elements

$$C_{ij} = \sum_{k=1}^{q} A_{ik} B_{kj} .$$

This can be interpreted in the standard pictorial form of taking a row of $\mathbf{A}$ times a column from $\mathbf{B}$. This only works if the number of columns of $\mathbf{A}$ is equal to the number of rows of $\mathbf{B}$. In practice, we will primarily be interested in square matrices, in which case *p*=*q*=*r* and thus the dimensions of the matrices will always match up. The important point is that matrix multiplication *does not commute.* This is easily proven using 2 by 2 matrices:

$$\mathbf{AB} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 4 & 3 \end{pmatrix}$$

$$\mathbf{BA} = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 4 \\ 0 & 3 \end{pmatrix}$$

Another important point is that the transpose of a product of matrices is given by the product of the transposes *in the reverse order*:

$$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T.$$

There are two special kinds of matrices: n by 1 matrices (which we call **column vectors**) and 1 by m matrices (which we call **row vectors**).

$$\begin{pmatrix} \\ \\ \\ \end{pmatrix} \quad \textit{and} \quad (\qquad)$$

We will use lowercase bold letters ($\mathbf{x}$, $\mathbf{y}$ etc.) to denote vectors. Column and row vectors are related through transposition; the transpose of a row vector is a column and vice-versa:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \end{pmatrix} \Rightarrow \mathbf{x}^T = \begin{pmatrix} x_1 & x_2 & \ldots \end{pmatrix}$$

As shown above, we will usually give only one index to the elements of a vector, since the other (either the row index for a column vector or the column index for a row vector) is always '1'. The product of a row vectors and a column vector is a 1 by 1 matrix – a number. This number is special; it is the **dot product** (or inner product) between the vectors, which we will write

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y} = \begin{pmatrix} x_1 & x_2 & \ldots \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \cdots \end{pmatrix} = \sum_i x_i y_i$$

The dot product tells us to what extent the two vectors point "in the same direction". Using the dot product, we can define the **norm** of a column vector:

$$|\mathbf{x}|^2 = \mathbf{x}^T \mathbf{x} = \begin{pmatrix} x_1 & x_2 & \ldots \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdots \end{pmatrix} = \sum_i x_i^2$$

In addition to the inner product of two vectors, which is always "row times column" we will also be interested in the **outer** (or **tensor**) **product** which involves the opposite ordering:

$$\mathbf{y} \otimes \mathbf{x} = \mathbf{y}\ \mathbf{x}^T = \begin{pmatrix} y_1 \\ y_2 \\ \dots \end{pmatrix} \begin{pmatrix} x_1 & x_2 & \dots \end{pmatrix} = \begin{pmatrix} y_1 x_1 & y_1 x_2 & y_1 x_3 \\ y_2 x_1 & y_2 x_2 & \dots \\ y_3 x_1 & \dots & \dots \end{pmatrix}$$

Note that instead of giving a number, this gives us a matrix. Many times, we will omit the "⊗"; in this case it is understood that for "column times row" we mean outer product.

We now define several terms that will be useful later:

A vector is **normalized** if $|\mathbf{x}|^2 = 1$. Given any un-normalized vector, one can normalize it by dividing by a constant, $c = |\mathbf{x}|$.

Two vectors are **orthogonal** if $\mathbf{x} \cdot \mathbf{y} = 0$. If the vectors are both orthogonal and normalized, they are said to be **orthonormal**.

A matrix is symmetric if $\mathbf{A}^T = \mathbf{A}$. It is anti-symmetric (or skew-symmetric) if $\mathbf{A}^T = -\mathbf{A}$.

A matrix is **orthogonal** if $\mathbf{A}^T \mathbf{A} = \mathbf{1}$, where $\mathbf{1}$ is the unit matrix (ones on the diagonal, zeros on the off-diagonal. For a square orthogonal matrix, $\mathbf{A}$, the transpose, $\mathbf{A}^T$, is also orthogonal.

If we treat each column of the matrix as a vector, saying the matrix is orthogonal is equivalent to saying that the columns are all *orthonormal*. To see this, begin with the fully indexed form of $\mathbf{A}^T \mathbf{A} = \mathbf{1}$:

$$\mathbf{A}^T \mathbf{A} = \mathbf{1} \quad \Rightarrow \quad \sum_k (A^T)_{ik} A_{kj} = \delta_{ij}$$

where we have introduced the **Kronecker delta**, δ_{ij}, which is defined to be one if i=j and zero otherwise. It is just a handy way of writing the unit matrix. Taking the transpose of $\mathbf{A}$ just swaps the order of the indices (cf. the definition of the transpose above) so:

$$\sum_k (A^T)_{ik} A_{kj} = \delta_{ij}$$
$$\Rightarrow \sum_k A_{ki} A_{kj} = \delta_{ij}$$

We now group the columns of $\mathbf{A}$ into vectors $\mathbf{a}_i$:

$$\mathbf{A} = (\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3 ...)$$

the vectors are given by:

$$\mathbf{a}_i \equiv \begin{pmatrix} A_{1i} \\ A_{2i} \\ A_{3i} \\ \dots \end{pmatrix}$$

We can then re-write $\mathbf{A}^T\mathbf{A} = \mathbf{1}$ in terms of the *dot products* of these vectors:

$$\sum_k A_{ki} A_{kj} = \delta_{ij}$$
$$\Rightarrow \mathbf{a}_i^T \cdot \mathbf{a}_j = \delta_{ij}$$

This last equation is just another way of saying the vectors are orthonormal: the dot product between any two different vectors (i≠j) is zero while the norm of each vector (i=j) is 1.

We will be interested in some fairly abstract vector spaces, and I have already hinted that these spaces will generally be infinite dimensional. How do we even define dimension in such a case? First we note that in any vector space, one can always come up with a (possibly infinite) set of vectors,

$$\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, ...\}$$

such that every *other* vector can be written as a linear combination of the vectors in $\mathcal{S}$:

$$\textit{any} \quad \mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + ...$$

$\mathcal{S}$ is then said to "span the space" we are studying. Now, there might be 'extra' vectors in $\mathcal{S}$. For example, we might be able to write $\mathbf{v}_1$ as a linear combination of $\mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4 ...$:

$$\mathbf{v}_1 = b_2\mathbf{v}_2 + b_3\mathbf{v}_3 + b_4\mathbf{v}_4 ...$$

In this case we could remove $\mathbf{v}_1$ from $\mathcal{S}$ and it would still "span the space". If we remove all these redundant vectors from $\mathcal{S}$ (so that no vector in $\mathcal{S}$ can be written as a linear combination of the others) then $\mathcal{S}$ becomes a 'linearly independent basis', or simply a **basis**. Typically, there will be many different sets, $\mathcal{S}$, that form a basis for a given space. However, it can be shown that for any well-behaved vector space, the *number* of vectors in any basis is always the same and this number is defined to be the **dimension** of the space. If all the vectors in the basis are also orthonormal (i.e. if $\mathbf{v}_i^T \cdot \mathbf{v}_j = \delta_{ij}$) then we have an **orthonormal basis** for the space. It turns out that any basis can be made orthonormal, and the algebra will always turn out to be simpler if we assume our basis is orthonormal. So we will usually assume it.

As an example, say we have an orthogonal basis, $\{\mathbf{v}_i\}$ and we want to find the expansion coefficients c_i for an arbitrary vector, $\mathbf{x}$:

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + ... = \sum_i c_i \mathbf{v}_i$$

If we take the inner product of $\mathbf{x}$ with $\mathbf{v}_j$ we find:

$$\mathbf{v}_j^T\mathbf{x} = c_1\mathbf{v}_j^T\mathbf{v}_1 + c_2\mathbf{v}_j^T\mathbf{v}_2 + c_3\mathbf{v}_j^T\mathbf{v}_3 + ... = \sum_i c_i \mathbf{v}_j^T\mathbf{v}_i$$

however, *if the basis is orthonormal* then $\mathbf{v}_j^T\mathbf{v}_i = \delta_{ij}$ which allows us to write:

$$\mathbf{v}_j^T\mathbf{x} = \sum_i c_i \mathbf{v}_j^T\mathbf{v}_i = \sum_i c_i \delta_{ij} = c_j$$

Thus, in an orthonormal basis, we can obtain a simple expression for the expansion of an arbitrary vector: $c_j = \mathbf{v}_j^T\mathbf{x}$.

The choice of a basis for the vector space is completely arbitrary; clearly the properties of the vector $\mathbf{x}$ cannot depend on the basis we choose to represent it. This fundamental realization leads to two important points. The first is that if we multiply any vector by the matrix $\mathbf{M} = \sum_j \mathbf{v}_j\mathbf{v}_j^T$, we get the same vector back again:

$$\mathbf{Mx} = \sum_j \mathbf{v}_j\mathbf{v}_j^T\mathbf{x} = \sum_j \mathbf{v}_j\left(\mathbf{v}_j^T\mathbf{x}\right) = \sum_j c_j\mathbf{v}_j = x$$

The matrix that always gives the same vector back after multiplication is the unit matrix. So we must have that

$$\sum_j \mathbf{v}_j\mathbf{v}_j^T = \mathbf{1}$$

and we say that a basis forms a "resolution of unity" or a "resolution of the identity".

The second point is that given two different bases ($\{\mathbf{v}_i\}$ and $\{\mathbf{u}_i\}$), we would like to be able to perform a **change of basis**. That is, given the expansion coefficients of $\mathbf{x}$ in one basis, we want to be able to directly obtain the expansion coefficients in another basis. The key object in this is the transformation matrix:

$$T_{ij} = \mathbf{u}_i^T \mathbf{v}_j = \mathbf{v}_j^T \mathbf{u}_i$$

This matrix is orthogonal:

$\{\mathbf{v}_i\}$ orthonormal

$$\mathbf{T}^T\mathbf{T} = \sum_k T_{ki}T_{kj} \quad = \quad \sum_k \mathbf{v}_i^T \mathbf{u}_k \mathbf{u}_k^T \mathbf{v}_j \quad = \quad \mathbf{v}_i^T \mathbf{v}_j = \delta_{ij}$$

$$\sum_j \mathbf{u}_j \mathbf{u}_j^T = \mathbf{1}$$

and it converts the expansion coefficients from the $\{\mathbf{v}_i\}$ basis to the $\{\mathbf{u}_i\}$ basis:

$$\sum_j \mathbf{u}_j \mathbf{u}_j^T = \mathbf{1}$$

$$\mathbf{x} = \sum_j c_j \mathbf{v}_j = \sum_k \mathbf{u}_k \mathbf{u}_k^T \sum_j c_j \mathbf{v}_j = \sum_k \mathbf{u}_k \sum_j c_j \mathbf{u}_k^T \mathbf{v}_j = \sum_k \mathbf{u}_k \sum_j c_j T_{kj}$$

Now, the expansion coefficients in the $\{\mathbf{u}_i\}$ basis are defined by:

$$\mathbf{x} = \sum_j b_j \mathbf{u}_j$$

and so we can quickly read of from the revious equation the equality:

$$b_k = \sum_j c_j T_{kj}$$

Or, in vector form:

$$\mathbf{b} = \mathbf{Tc}.$$

So we see that *orthogonal matrices allow us to transform from one basis to another*.

Another important property of a matrix is that it has **eigenvalues** and **eigenvectors**. These special vectors satisfy:

$$\mathbf{M}\mathbf{x}_i = m_i \mathbf{x}_i$$

The numbers m_i and the vectors $\mathbf{x}_i$ are the eigenvectors of the matrix. Actually, the $\mathbf{x}_i$ are the right eigenvectors of the system. A matrix also has left eigenvectors:

$$\mathbf{y}_i\mathbf{M} = m_i\mathbf{y}_i$$

and in general these two sets of vectors can be different. However, in the very important case that $\mathbf{M}$ is *symmetric*, the left and right vectors are the same. There are two more important features of symmetric matrices in this regard: 1) their eigenvalues are always *real* and 2) their eigenvectors can be chosen to form an orthonormal basis. These are very nice properties. The only ambiguity is in the case of degenerate eigenvalues; in this case some of the degenerate eigenvectors may not be orthogonal to one another, but one can always choose an appropriate linear combination of the degenerate eigenvectors to make them orthonormal. Note that a linear combination of two degenerate eigenvectors is, again an eigenvector. To see this, assume $\mathbf{x}_1$ and $\mathbf{x}_2$ have the same eigenvalue (m). Then:

$$\mathbf{M}(a\mathbf{x}_1 + b\mathbf{x}_2) = a\mathbf{M}\mathbf{x}_1 + b\mathbf{M}\mathbf{x}_2 = am\mathbf{x}_1 + bm\mathbf{x}_2 = m(a\mathbf{x}_1 + b\mathbf{x}_2)$$

A symmetric matrix can be **diagonalized** in terms of its eigenvalues and eigenvectors. Define the orthogonal matrix that has the eigenvectors as its columns:

$$\mathbf{X} = (\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3 \quad \mathbf{x}_4 ...)$$

and a diagonal matrix $\mathbf{D}$ that has the eigenvalues on the diagonal and zeros elsewhere. Then

$$\mathbf{M} = \mathbf{X}\mathbf{D}\mathbf{X}^T$$

To see this, we merely insert unity on both sides of $\mathbf{M}$:

x$_j$ is an eigenvector

$$\sum_i \mathbf{x}_i\mathbf{x}_i^T\mathbf{M}\sum_j \mathbf{x}_j\mathbf{x}_j^T = \sum_i \mathbf{x}_i\mathbf{x}_i^T\sum_j \mathbf{M}\mathbf{x}_j\mathbf{x}_j^T = \sum_i \mathbf{x}_i\mathbf{x}_i^T\sum_j m_j\mathbf{x}_j\mathbf{x}_j^T$$

$$= \sum_i \mathbf{x}_i\sum_j m_j\mathbf{x}_i^T\mathbf{x}_j\mathbf{x}_j^T = \sum_i \mathbf{x}_i\sum_j m_j\delta_{ij}\mathbf{x}_j^T = \sum_i \mathbf{x}_i m_i\mathbf{x}_i^T$$

Eigenvectors are Orthonormal

which is just the right hand side of the equation above.

Finally, we will also be interested in a **function** of a matrix. This is defined formally by the power series. For example:

$$e^{\mathbf{M}} = \mathbf{1} + \mathbf{M} + \tfrac{1}{2!}\mathbf{M}\mathbf{M} + \tfrac{1}{3!}\mathbf{M}\mathbf{M}\mathbf{M} + ...$$

For a symmetric matrix, this can be greatly simplified by going to the diagonal representation:

$$e^{\mathbf{M}} = \mathbf{1} + \mathbf{XDX}^T + \tfrac{1}{2!}\mathbf{XD}\cancel{\mathbf{X}^T\mathbf{X}}\mathbf{DX}^T + \tfrac{1}{3!}\mathbf{XD}\cancel{\mathbf{X}^T\mathbf{X}}\mathbf{D}\cancel{\mathbf{X}^T\mathbf{X}}\mathbf{DX}^T + \ldots$$

$$\Rightarrow e^{\mathbf{M}} = \mathbf{1} + \mathbf{XDX}^T + \tfrac{1}{2!}\mathbf{X}\overset{1}{\mathbf{D}}\mathbf{DX}^T + \tfrac{1}{3!}\mathbf{X}\overset{1}{\mathbf{D}}\overset{1}{\mathbf{D}}\mathbf{DX}^T + \ldots = \mathbf{X}e^{\mathbf{D}}\mathbf{X}^T$$

where we have made repeated use of the fact that the eigenvectors are orthonormal, so $\mathbf{X}^T\mathbf{X} = \mathbf{1}$. The power of this last expression is that it allows us to re-phrase the evaluation of the exponential of an arbitrary matrix in terms of the exponential of a *diagonal* matrix. The exponential of a diagonal matrix is easy (one just exponentiates the diagonal elements). The same transformation works for any function: $f(\mathbf{M}) = \mathbf{X}f(\mathbf{D})\mathbf{X}^T$.