Matrix Chapter 10 (Margenau and Murphy) PhL start: 12.5.04
finish: 2.12.04

Motivation: My matrix theory information seems to be scattered in little bits and pieces, I don't have any good book that covers the things I am interested in. Maybe those lost red books had this stuff. Maybe I can find a book in the library. Maybe the new Kazdan I ordered. I do have this M&M chapter however, it is about 30 pages long. Let's see if we can at least find out WHAT is IN this chapter? I will prove things along the way. // In retrospect, it did provide the foundation for what I want to know.

[Note added: I have several other matrix documents now that should accompany these notes. Those other documents really have more important stuff, but some things are only located in *this* document. For that reason, I went through and numbered all the theorems in these notes so they can be referred to.]

****

[Warning: We use a and A to mean the same thing. Also det(A) = |A|.]

10.2 Determinants. Applies only to a square NxN matrix. A definition is given which is a verbal form of this formula: (sum 1 to N implied on all repeated indices)

$$\det a = \varepsilon_{\ell mn....q} \; a_{1\ell}\, a_{2m}\, a_{3n}.....a_{Nq}$$

This is a sum of terms where each term consists of a product of N matrix elements taken one per row, as the first index shows. The ε symbol is +1 for $\varepsilon_{123....N}$ and changes to -1 if you swap any pair of labels. Thus, for an even number of swaps it is still +1, but for an odd number, it is -1. If any two indices take the same value, ε = 0. Thus, as you write out one of the non-zero terms, you have to take each of the elements in a different column. So in your selection of $a_{1\ell}$ you pick some column ℓ, there are n choices, Then for your choice of a_{2m} you must pick a different column, so (n-1) choices. Therefore, there are really only N! terms in this sum, just as they say (not N^2) . One term in the sum is the product of all the diagonal elements and that term has a + sign.

You can write the same determinant in this form:

$$\det a = \varepsilon_{\ell mn....q} \; a_{\ell 1}\, a_{m2}\, a_{n3}.....a_{qN}$$

Again, each non-zero term has all its factors in different rows and different columns. Since the diagonal element has the same sign as our previous writing, these two sums must be the same, because they have exactly the same terms.

Theorem 2Z. det(a) = det(a^T), so determinant is the same if you swap rows and columns.

Proof: $\det a = \varepsilon_{\ell mn....q} \; a_{1\ell}\, a_{2m}\, a_{3n}.....a_{Nq}$ and $\det a^T = \varepsilon_{\ell mn....q} \; a_{\ell 1}\, a_{m2}\, a_{n3}.....a_{qN}$ where we have used the fact that $(a^T)_{jk} = a_{kj}$. But this last ε expression is just the second form of det(a) shown above, QED.

Theorem 2A: If a row = 0, then det = 0. // same for col

Proof: all terms in det must have a factor from that row, but all its elements are zero, hence det = 0.

Lemma 2B for next Theorem: $\varepsilon_{\ell mn....q}M_{\ell mn....q} = 0$ if M is symmetric in *any* pair of indices

Proof: If you swap the symmetric pair of indices on M, say $\ell \leftrightarrow$ m, M does not change. Then swap the names of the two dummy summation indices $\ell \leftrightarrow$ m. If you then swap $\ell \leftrightarrow$ m on ε, you pick up a minus sign, and you conclude that (what you started with) = - (what you started with), so it must be 0.

Theorem 2C: If one row is a multiple of another, then det = 0. // same for col

Proof: Suppose $a_{3n} = ka_{2n}$ for all n (row 3 is multiple of row 2). Then det a = k $\varepsilon_{\ell mn....q}$ $a_{1\ell}$ a_{2m} $a_{2n}.....a_{Nq}$. But now $M_{\ell mn....q} = k\, a_{1\ell}\, a_{2m}\, a_{2n}.....a_{Nq}$ is symmetric for n $\leftrightarrow$ m, so det = 0 by above lemma.

Theorem 2D: Add a multiple of one row to another row, det does not change.

Proof: Suppose $a_{3n} \rightarrow a_{3n} + ka_{2n}$ (add multiple of row 2 to row 3). The second term this creates in the det vanishes by the previous theorem, so det is unchanged. Same if you did $_{3n} \rightarrow a_{3n} + ka_2 + k'a_{4n}$ where you now add multiples of two rows, and so on.

Lemma 2E for next Theorem: If you swap two indices on M, sign of $\varepsilon_{\ell mn....q}M_{\ell mn....q}$ changes.

Proof: Let $M'_{\ell nm....q} \equiv M_{\ell mn....q}$ where we did n $\leftrightarrow$ m. As in previous lemma, swap on ε to get a - sign, then swap dummy sum indices n $\leftrightarrow$ m and you get that $\varepsilon_{\ell mn....q}\, M'_{\ell mn....q} = -\varepsilon_{\ell mn....q}M_{\ell mn....q}$

Theorem 2F: Swap two columns and sign of det changes.

Proof: Follows directly from previous lemma where $M_{\ell mn....q} = a_{1\ell}\, a_{2m}\, a_{3n}.....a_{Nq}$

Theorem 2G: Multiply a column by k, you then multiply det by k.

Proof: obvious from above def of det, the k just comes out front.

Theorem 2H. The determinant of an **upper triangular matrix** is the product of the diagonal elements.

Proof: Draw a picture of such a matrix. The only term in the determinant sum will have to contain the first diagonal element because it is the ONLY element in the first column. As we move to the second column, we can only take the 2nd element, since this is the only non-zero element not in the first row. Continuing, we see that the only term of the N! terms in the determinant is the product of the diagonal elements.

****

10.3 Minors and Cofactors

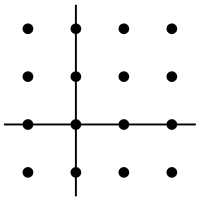
Define: **minor(a_{pq})** = det of the N-1 x N-1 matrix you get by crossing out row p and column q.

Define: **cof(a_{pq})** = $(-1)^{p+q}$ minor(a_{pq}).

Theorem 3A: $\det(A) = \Sigma_q\, a_{pq}\ \mathrm{cof}(a_{pq})$ [across row p] = $\Sigma_p\, a_{pq}\ \mathrm{cof}(a_{pq})$ [down column q] (10-2)

It takes a little work to prove this theorem, and an example first helps. One of the problems is that the submatrices you deal with have non-standard numbering because you have crossed out things.

Example: consider a 4x4 matrix where we cross out row 3 and column 2.



The determinant of a 4x4 matrix A is given by

$$\det(A) = \varepsilon_{ijk\ell}\ a_{1i}\ a_{2j}\ a_{3k}\ a_{4\ell} \qquad \text{// 1,2,3,4 are row indices}$$

with 4 sums implied. The minor shown above can then be written in this way (as we shall explain)

$$\text{minor}(a_{32}) = -\ \varepsilon_{ij2\ell}\ a_{1i}\ a_{2j}\ a_{4\ell} \qquad \text{// } a_{3k} \text{ is missing, and we set k = 2}$$

with an implied triple sum. The 2 index on the ε removes all terms in the three sums in which some matrix element column index is 2, so we really are summing over the right data so that this is the determinant of the submatrix obtained by crossing out column 2 and row 3. The only question is the overall sign. If we look at the "diagonal" term ("diagonal" of the 3x3 submatrix) we get i=1,j=3, ℓ=4 for column indices and that term is then $-\varepsilon_{1324}\ a_{11}\ a_{23}\ a_{44}$. But this is equal to $+\ \varepsilon_{1234}\ a_{11}\ a_{23}\ a_{44}$ $= +\ a_{11}\ a_{23}\ a_{44}$ and that is the correct + sign for the diagonal term of a determinant sum.

In this example, notice that we set the "3rd index [k] (p)" to "2 (q)" on the ε tensor. The reason we did this is as follows: (1) since we crossed out the third ROW, we killed off the a_{3k} factor in the product, and so the k index is no longer present, and k was the third (p_{th}) index on ε. (2) We forced this index to be q=2 on the ε tensor in order to remove terms in the implied sums that involved COLUMN 2 = q.

We then incurred a - sign in order to shift the "2" to the second position from the third position on ε.

In the general case of ROW p and COLUMN q, at the p^{th} index position of the ε we will have the fixed number q. In order to move this fixed number q to its "standard index position", we will have to do some swaps. We will do | p - q | swaps to get it into position, just as we did | 3 - 2 | = 1 swap in the example. This will cause a sign factor of $(-1)^{|p-q|}$. If | p - q | is odd, then p+q is odd, since | p - q | = p + q - 2*min(p,q). Therefore, our sign factor can be written $(-1)^{p+q}$, which is the same sign in the definition of the cofactor. Therefore we have shown that

$\text{cof}(a_{32}) = \varepsilon_{ij2\ell}\, a_{1i}\, a_{2j}\, a_{4\ell}$ // example

$\text{cof}(a_{pq}) = \varepsilon_{ij.....q.....\ell}\, a_{1i}\, a_{2j} ..... a_{p-1x}\, \cancel{(a_{py})}\, a_{p+1z} .....$ // in general

where we have now tried to write a very general result. It has the number q in the p^{th} ε subscript position, and the a_{py} factor is *missing* from the product. (we show it "struck out").

Now comes the final result we are seeking: suppose we multiply the above cofactor by the missing factor a_{pq} and then sum on q , leaving p fixed. What do we get?

$$\Sigma_q\, a_{pq}\ \text{cof}(a_{pq}) = \varepsilon_{ij.....q.....\ell}\, a_{1i}\, a_{2j} ..... a_{p-1x}\, a_{pq}\, a_{p+1z} ..... = \det(A)$$

The reason is that now q is like any other summation index, and we slide the a_{pq} into its proper position in the product of factors, and we then recognize our definition of det(A). In this last formula on the left side, we are selecting a particular fixed row p, and *we work across that row* in our q sum. It can be any row.

Since the det(a) = det(a^T), we can rewrite our result $\det(A) = \Sigma_q\, a_{pq}\, \text{cof}(a_{pq})$ as $\Sigma_q\, a_{qp}\ \text{cof}(a_{qp})$. Now we swap the indices (just labels) to get $\det(A) = \Sigma_p\, a_{pq}\, \text{cof}(a_{pq})$ just to get the same result as earlier but now summed on p for fixed q. This says we can do the cofactor evaluation by picking a column q and *working down that column*. It can be any column.

These two results appear as equation (10-2) in M&M, with this identification:

$A_{pq} = a_{pq}$ = matrix element $A^{pq} = \text{cof}(a_{pq})$ // upper indices on A

Theorem 3B: $\Sigma_q\, a_{pq}\ \text{cof}(a_{p'q}) = 0$ when $p \neq p'$. Also $\Sigma_p\, a_{pq}\ \text{cof}(a_{pq'}) = 0$ when $q \neq q'$

Proof: Start with a matrix A and take the data from row p and overwrite row p' with it. Call this new matrix A', and note that it has two identical rows and therefore detA' = 0 by an earlier theorem. Now since in general we know that $\det A = \Sigma_q\, a_{pq}\ \text{cof}(a_{pq})$, we can write $\det A' = \Sigma_q\, a_{pq}\ \text{cof}(a_{p'q})$ since $a_{pq} = a_{p'q}$ by construction of matrix A' (row p and row p' are the same). Therefore $\Sigma_q\, a_{pq}\ \text{cof}(a_{p'q}) = 0$, QED.

A similar argument applies to the other form of the det, so we get also that $\Sigma_p\, a_{pq}\ \text{cof}(a_{pq'}) = 0$ when $q \neq q'$. These two equations appear as (10-3) in M&M.

Now we can combine these two sets of equations together: (replace a with A)

$\Sigma_p\, A_{pq}\ \text{cof}(A_{pq'}) = \delta_{q,q'}\, \det(A)$ and $\Sigma_q\, A_{pq}\ \text{cof}(A_{p'q}) = \delta_{p,p'}\, \det(A)$

Notational issue: M&M refer to the cofactor matrix as $A^{pq} = \text{cof}(A_{pq})$, putting superscripts on A. Suppose we were to just write $C_{pq} = \text{cof}(A_{pq})$. We might call this C thing the "cofactor matrix". Then clearly $(C^T)_{pq} = \text{cof}(A_{qp})$. It is little hard to write this without defining a symbol like C. When we write $\text{cof}(A_{pq})$ we mean a number which is "the cofactor of matrix element A_{pq}". Maybe better to say $\text{cof}(A)_{pq}$ to imply that cof(A) is a matrix, and that a matrix element is given by $\text{cof}(A)_{pq} = \text{cof}(A_{pq})$.

Now let's rewrite our equations above:

$\Sigma_p \text{ cof}(A)^T_{q'p} A_{pq} = \delta_{q,q'} \det(A)$ and $\Sigma_q A_{pq} \text{ cof}(A)^T_{qp'} = \delta_{p,p'} \det(A)$

Now the summations are in proper matrix form, and we can write these again as:

$$\text{cof}(A)^T A = \det(A)\, E \qquad \text{and} \qquad A\, \text{cof}(A)^T = \det(A)\, E$$

where E is the identify matrix. This shows that

$$A^{-1} = (1/\det(A))\, \text{cof}(A)^T$$

and we have derived an explicit formula for the inverse of n arbitrary NxN matrix A. The formula of course fails if det(A) = 0. We now just use an M&M definition of this transposed cofactor matrix,

Definition: The **adjoint matrix** of A, written $A^\wedge$, is the transpose of the cofactor matrix, $A^\wedge \equiv \text{cof}(A)^T$. Note that the word "adjoint" means different things to different authors, so beware. This is what it means to M&M.

With this definition, we get $AA^\wedge = A^\wedge A = \det(A)\, E$. So we have proven the following:

Theorem 3C: For any square matrix A, $A^{-1} = (1/\det(A))\, \text{cof}(A)^T = (1/\det(A))A^\wedge$, provided $\det(A) \neq 0$, which is to say, provided matrix A is not singular.

****

10.4 Multiplication and Differentiation of Determinants.

Theorem 4A: det(AB) = det(A) det(B)

M&M allow for the product to be summed in any ONE of four index ways, I always think in terms of the first and most standard meaning of the product C = AB. Proof is not trivial.

Proof: I think I have a pretty clever proof here. I will just show 3 indices for this proof, the generalization is obvious. The RHS of the above is this:

$$\left(\Sigma_{ijk}\, \varepsilon_{ijk}\, A_{1i} A_{2j} A_{3k} \right) \left(\Sigma_{i'j'k'}\, \varepsilon_{i'j'k'}\, A_{1i'} A_{2j'} A_{3k'} \right) = \det A \det B$$

In the second expression, as usual we order the rows 1,2,3... . But we could choose any order we like, as long as we sign-correct with an ε. For example, we could say

$$\left(\Sigma_{ijk}\, \varepsilon_{ijk}\, A_{1i} A_{2j} A_{3k} \right) \left(\Sigma_{i'j'k'}\, \varepsilon_{i'j'k'}\, B_{Ii'} B_{Jj'} B_{Kk'}\, \varepsilon_{IJK} \right)$$

where the fixed integers I,J,K can be *any* permutation of 1,2,3. Suppose, for each term in the sum on the left, we "customize" our choice of IJK to be equal to ijk. Here then is what the ijk^{th} term in the left sum would look like:

$$\varepsilon_{ijk}\ A_{1i}A_{2j}A_{3k}\ \ x\ \ \ (\ \Sigma_{i'j'k'}\ \varepsilon_{i'j'k'}\ B_{ii'}B_{jj'}B_{kk'}\ \varepsilon_{ijk})$$

So now we reapply the left side sum after doing this customization to get:

$$\Sigma_{ijk}\ \{\ \ \varepsilon_{ijk}\ A_{1i}A_{2j}A_{3k}\ \ (\ \Sigma_{i'j'k'}\ \varepsilon_{i'j'k'}\ B_{ii'}B_{jj'}B_{kk'}\ \varepsilon_{ijk}\ \}\ \}\ \ \ = \text{detA*detB}$$

We know that $\varepsilon_{ijk}^2 = 1$ so those factors go. Then slide the primed sums and ε to the left:

$$=\ \ \Sigma_{i'j'k'}\varepsilon_{i'j'k'}\ [\ \Sigma_{ijk}\ \ (A_{1i}B_{ii'})\ (A_{2j}B_{jj'})(A_{3k}B_{kk'})\ ]$$

where we have grouped the factors to prepare for the next step. Now we use this fact:

$$\Sigma_{ijk}\ \ X_iY_jZ_k\ \ = \Sigma_{ij}\ \ X_iY_j\ (\Sigma_kZ_k\ )\ =\ \Sigma_i\ X_i\ (\Sigma_kZ_k\ )(\Sigma_j\ \ Y_j) = (\Sigma_kZ_k\ )(\Sigma_j\ \ Y_j)(\Sigma_i\ X_i)$$

and we then get

$$= \Sigma_{i'j'k'}\varepsilon_{i'j'k'}\ [\ \{\ \Sigma_i(A_{1i}B_{ii'})\}\ \{\Sigma_j(A_{2j}B_{jj'})\}\ \{\Sigma_k(A_{3k}B_{kk'})\}\ ]$$

$$=\ \Sigma_{i'j'k'}\varepsilon_{i'j'k'}\ \ C_{1i'}\ \ C_{2j'}C_{3k'}\ \ \ = \text{det(C)}\qquad\qquad\qquad\qquad \text{QED.}$$

Theorem 4B: $\partial\ |A|/\partial a_{pq} = \text{cof}(a_{pq})$

Proof: Recall from above that

$$|A| = \Sigma_q\ a_{pq}\ \ \text{cof}(a_{pq})\ \ =\ \ \varepsilon_{ij.....q.....\ell}\ \ a_{1i}\ a_{2j}\ .....\ a_{p-1x}\ a_{pq}\ a_{p+1z}\ .....$$

In the summation $\Sigma_q\ a_{pq}\ \ \text{cof}(a_{pq})$, the matrix element a_{pq} appears ONLY where you see it explicitly. Thus, the partial derivative wrt it gives the result claimed. This appears on top of page 305 or M&M.

****

10.5 Preliminary Remarks on Matrices

Authors describe their notation conventions, I agree with all.

Def: A square matrix A is **singular** if det(A) = 0. All rectangular matrices are defined as singular, where rectangular means non-square.

Def: **Rank of A**: If det(A) ≠ 0, A has full rank N. If det(A) = 0 but if ALL det(minors) ≠ 0, then rank is N-1. And so on. Rank applies only to a square matrix.

Therefore, if A is non-singular, rank = N, but if A is singular, rank < N.

Theorem 5A: If rank A = r, then we know two things:

(1) *all* RxR sub-determinants of A for R > r vanish.

(2) *at least one* rxr sub-determinant does *not* vanish

****

10.6 Combinations of Matrices

This section defines matrices as being **conformable** if you can multiply them together. Matrices might not **commute**. You can make a **dot product** or a new matrix from two vectors, combining them in the two ways we know. You can work with **partitioned** matrices, and you can make a **direct product** matrix. M&M will use [x] for **row vector** and {x} for **column vector**.

****

10.7 Special Matrices

We have the O and E identity matrices, and we can have diagonal matrices.

Theorem 7A: If [A,D] = 0 where D is diagonal, A_{ij} [D_i - D_j] = 0, so the only possible *non-zero* off-diagonal elements of A_{ij} are when $D_i = D_j$. For example, if $D_2 = D_3$, then you can have $A_{23} \neq 0$. Thus, if all diagonal elements are *different*, A must be diagonal.

Structurally, if D has some identical elements and you order them properly, then matrix A must be diagonal in the sense of boxes on the diagonal. We write A = diag(B_1,B_2 .....)

Corollary: It is NOT true that AD = DA for an arbitrary matrix A when D is diagonal. A must be in block diagonal form consistent with the multiplicity of the diagonal values of D.

The **trace** is defined (= spur), and we have

Theorem 7B: Tr(AB) = Tr(BA), whether or not the matrices commute (Proof: $A_{ij}B_{ji}$ = reverse)

Theorem 7C: Tr(AxB) = Tr(A)Tr(B). (Proof: from definition of the cross product.)

def**: Transpose** Matrix: $A^{\sim}_{k\ell} = A_{\ell k}$ (I usually say A^T for transpose, they use twiddle over the A)

def**: Adjoint** Matrix: $A^{\wedge}_{pq} = \text{cof}(a_{pq})^T$ (I don't ever use this matrix, beware the name, see above.)

Theorem 7D: $A^{\wedge}A = E * \det(A) = AA^{\wedge}$

Theorem 7E: If A is non-singular, then $A^{-1} = A^{\wedge}/|A|$.

See end of section 10.3 notes above for proofs and comments on these two theorems.

Fact: If you transpose or invert a product of matrices, you have to change the order.

def: **Complex Conjugate Matrix:** $A^*_{ij} = a_{ij}^*$

def: **Associate Matrix**: $A^{\dagger} = A^{*\sim} = A^{*T}$

Warning: In Stakgold, the matrix $A^{\dagger}$ is written A*, and is called the adjoint of A. This is a double confusion because (1) adjoint means cof^{T} to M&M, and (2) * means C.C. to M&M. Stakgold then must use the overbar for C.C. In Quantum Mechanics, $A^{\dagger}$ is usually called the "Hermitian conjugate" of A.

Finally we get to the famous Table 1 where all the famous matrix types are shown.

symmetric, **orthogonal**, **real**, **Hermitian**, **unitary** and "**skew**" versions of some of these.

****

10.8 Real Linear Vector Space

An interesting way to write the vector cross product **c** = **a** x **b** is given. You have to make a 3x3 matrix A that is skew-symmetric, they you can write **c** = A**b**.

Theorem 8A: If basis vectors are linearly dependent, the **Gram Determinant** shown on page 312 vanishes. Conversely, if it does not vanish, then you have a basis by using the rows or columns.

Proof: If not dependent, then in the entire first column, replace with sum of others, $u_1 = \Sigma\, \alpha_i\, u_i$. Then you can take a multiple $-\alpha_2$ of the second column and add it to the first column to kill off the corresponding terms, and you do this for all other columns, and you end up with the left column being all 0! This makes the det = 0.

A good description of the **Gram-Schmidt Orthogonalization Process** is presented, very clear! Just a mechanical sequence of removing projections one at a time.

****

10.9 Linear Equations.

It seems to me that an important subject has been omitted at this point. The system A**x** = **y** represents N equations in N unknowns. After all, we are talking about equations in this section!

Def: A set of vectors $\mathbf{v}_i$ is **linearly independent** in E^n if *none* of them can be written as a lincom of some other ones.

Def: A set of vectors $\mathbf{v}_i$ is **linearly dependent** if the set is *not* linearly independent. In this case, at least one of the vectors must be writeable as a lincom of some others. If none were so writeable, then you would have linear independence. So it is an either/or situation with a set of vectors.

Example: The unit vectors $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are linearly independent because you cannot write $\hat{\mathbf{x}} = \alpha\, \hat{\mathbf{y}}$ no matter what value of α you try, including $\alpha = 0$.

Example: The unit vectors $\mathbf{0}$ and $\hat{\mathbf{y}}$ are linearly *dependent* because you can write $\mathbf{0} = \alpha\, \hat{\mathbf{y}}$ with $\alpha = 0$.

For any set of vectors, if the 0 vector is in the set, then the set is linearly dependent by our definition.

Phil Theorem 9.1: (General E^n) In E^n if the vectors $\mathbf{v_i}$ are linearly independent, then $\Sigma\, x_i\, \mathbf{v_i} = 0 \Rightarrow x_i = 0$. The converse is also true: $\Sigma\, x_i\, \mathbf{v_i} = 0 \Rightarrow x_i = 0 \;\Rightarrow$ the vectors $\mathbf{v_i}$ are linearly independent.

Proof: Start with the first claim. Suppose $x_i \neq 0$ for some i. Then we can solve $\Sigma\, x_i\, \mathbf{v_i} = 0$ for at least one $\mathbf{v_j} = \Sigma$ (other $\mathbf{v_i}$). But this says that the vectors are linearly dependent. So if the $\mathbf{v_i}$ are linearly independent, then $\Sigma\, x_i\, \mathbf{v_i} = 0 \Rightarrow x_i = 0$. Going in the converse direction, if $\Sigma\, x_i\, \mathbf{v_i} = 0 \Rightarrow x_i = 0$, then we cannot write any $\mathbf{v_j} = \Sigma$ (other $\mathbf{v_i}$). If we could, then that could give an equation of the form $\Sigma\, x_i\, \mathbf{v_i} = 0$ with some $x_i \neq 0$, which we have assumed cannot be done. Therefore, by our contrapositive above, get lin indep.

Interpretation of rows or columns of A as vectors: Consider a matrix A with N rows and M columns. Each row has M elements, so we can think of each row j as being a vector $\mathbf{R_j}$ in E^M. Similarly, since each column has N elements, so we can think of each column k as being a vector $\mathbf{C_k}$ in E^N. The connections are: $(\mathbf{R_j})_k = A_{jk}$ and $(\mathbf{C_k})_j = A_{jk}$. Notice that in the row interpretation the indices are not reversed, but in the column case they are. (The first index of A is the row index.)

Def: **linear independence of rows or columns** of matrix A. There is nothing to define here. We have shown above that the rows or columns can be regarded as vectors in a space, and vectors in a space can of course be linearly independent or not. If follows that if rows are linearly dependent, we can write at least one row as a lincom of other rows, and the same statement applies to columns.

Phil Theorem 9.2. If the N columns of a matrix A (possibly not square) are linearly independent, then $Ax=0 \Rightarrow x=0$. The converse is also true.

Proof: We can write $Ax=0$ as $\Sigma A_{jk} x_k = 0$ (j=1..N). We can interpret $A_{jk} = (\mathbf{C_k})_j$ as elements of the vector $\mathbf{C_k}$ where $\mathbf{C_k}$ is the k^{th} column of the matrix. Thus we have $Ax=0 \Rightarrow \Sigma A_{jk} x_k = 0 \Rightarrow \Sigma x_k\, (\mathbf{C_k})_j = 0$. If we now write all j of these equations, we can think of having one vector equation $\Sigma x_k\, (\mathbf{C_k}) = \mathbf{0}$ where now the 0 on the right is a vector 0. Since the columns are assumed to be independent, we get that x = 0 according to Phil Theorem 9.1. As for the converse, if $Ax=0 \Rightarrow x=0$, then $\Sigma x_k\, (\mathbf{C_k}) = \mathbf{0} \Rightarrow x = 0$ and by the converse of Theorem 9.1, we conclude that the $\mathbf{C_k}$ are linearly independent.

Phil Theorem 9.3: If the rows *or* cols of square matrix A are linearly dependent, then det(A) = 0.
Contrapositive: If $\det(A) \neq 0$, then the rows *and* cols of A are linearly independent.

Proof: If the cols are linearly dependent, we think of these cols as vectors in E^N and we are saying that we can write one col as a lincom of the other cols, so we do not have a basis. In this case, we know that we can do elementary operations on the determinant to clear out the col that we have written as a lincom of others. For example, suppose there are three columns $\mathbf{a_1}\ \mathbf{a_2}\ \mathbf{a_3}$. Suppose $\mathbf{a_1} = 3\mathbf{a_2} + 5\mathbf{a_3}$. Then in the

determinant we replace to get | $3\mathbf{a}_2 + 5\mathbf{a}_3$ $\mathbf{a}_2$ $\mathbf{a}_3$ |. But then we take -3 times the middle column and -5 times the right column and add these to the first column to clear it out, and we get detA = 0. The same argument can be given by regarding the rows as vectors instead of the columns.

Phil Theorem 9.4. If det(A) ≠ 0, then [Ax=0 ⇒ x=0].

Proof: If det(A) ≠ 0, then we can construct A^{-1} and apply to both sides of Ax=0 to get x = 0.

Phil Theorem 9.5. If the columns or rows of A are linearly independent, then det(A) ≠ 0.
Contrapositive: if detA = 0, then columns and rows are linearly dependent.

This is not obvious to me, but I think I can prove it with an algorithm. If the columns are linearly independent, then none can be the zero vector as noted above, so each column contains at least one non-zero number. So let's find a row that has a non-zero number in the left column and move this row to the top of the matrix. At most, this can only change the sign of detA since we may swap two rows to do this. Now use the usual Gauss elimination idea and clear out the entire first column below the top value.

Now at least one number in the second column must be non-zero. If there were only one non-zero number and if that were in the first row, then the first two columns would be multiples, and then detA = 0 contradicting our assumption. Therefore there must be a non-zero number in column 2 somewhere other than in the first row. Do a row swap if necessary to get this number (or any other non-zero number) into the second row. Now use that point as the pivot and use Gauss to clear out the rest of the second column.

Continue in this manner until we have converted A into a upper triangular matrix where all of the diagonal elements are non-zero. By Theorem 2H, the determinant of this matrix is the product of these diagonal elements, and is therefore not 0. But apart from a possible sign, this is also det(A), so we have proven then that detA ≠ 0 if the columns are linearly independent.

We can repeat the entire proof above using rows instead of columns, nothing changes in the argument.

Phil Theorem 9.6. Linear dependence of the columns ⇔ Linear dependence of the rows.
Contrapositive: Linear independence of the columns ⇔ Linear independence of the rows.

Proof: If the cols OR rows are linearly dependent, then detA=0 by Theorem 9.3. But then by the contrapositive of Theorem 9.5, the columns AND rows are linearly dependent. QED.

Phil Summary Theorem 9.7: By combinations of the above theorems, we can show that:

The following items all mean the same thing: (any one implies all the others)
(1) detA = 0, A is singular
(2) rows of A are linearly dependent
(3) columns of A are linearly dependent
(4) Ax = 0 does not imply that x=0, there can be non-zero solutions
(5) A^{-1} does not exist

The following items all mean the same thing: (any one implies all the others)
(1) detA ≠ 0, A is non-singular

(2) rows of A are linearly independent
(3) columns of A are linearly independent
(4) $Ax = 0 \Rightarrow x=0$
(5) A^{-1} exists

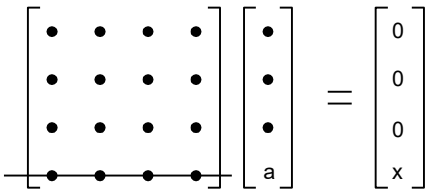
Fact: We can associate "**a set of N homogeneous equations in M unknowns**" with the matrix equation $Ax = 0$ where A has N rows and M columns. We can then associate each equation of this set with a specific row of A. Since each row can be thought of as a vector in space E^M (where M = number of columns of A), we identify each equation with such a row vector. The numbers in the row vector are the coefficients of our equation. We are trying here to stress this idea: equation = row = vector

For example, if we say that the *equations* of our set are linearly independent, we usually mean that we cannot write any one equation as a lincom of the other equations. Clearly this means the we cannot write any row as a lincom of the other rows. In the vector interpretation, we are then just saying that the row vectors form a basis in E^M.

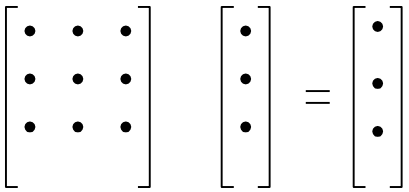
In the case that M = N we have a square matrix. In this situation, according to Phil Theorem 9.3, if $\det A \neq 0$, then we know that both the rows and columns are linearly independent in E^N. In this case, we know that our set of equations Ax=0 is linearly independent. In this case, Ax=0 has one solution, x=0.

What happens if detA = 0 (which still implies a square matrix) ? In this case, consider the contrapositive of the first statement in Phil Theorem 9.3. This first statement logically says $A+B \Rightarrow C$, so the contrapositive says $!C \Rightarrow !(A+B) = !A \bullet !B$. Thus, if $\det A \neq 0$, we know that the rows are not independent *and* that the columns are not independent. If the rows are not independent, then we know that the set of equations is not linearly independent, so at least one equation can be written as a sum of others. Since this equation is not then a distinct equation, so to speak, we really have N-1 equations in N unknowns. It could be that more than one of the equations in the set can be written in terms of a smaller set. If there are two such redundant equations, then we really have N-2 equations in N unknowns. In any case, with too few equations, there are going to be many solutions to Ax = 0 besides x = 0.

Exercise: Suppose we start with a 4x4 matrix equation Ax = 0, but we find that detA = 0. We decide that the last row of the matrix is a redundant equation, so we cross it out. We then have this situation:

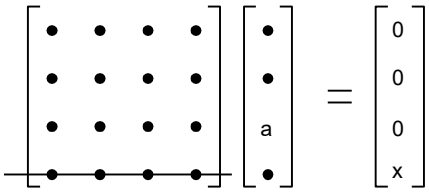


We then have 3 equations in 4 unknowns $x_1 \rightarrow x_4$. Suppose we then just select x_4 = a, some arbitrary value. Since we have removed the last equation, we don't care about the bottom 0 in the RHS vector. We have now reduced the problem to a 3x3 matrix equation of the form Bx' = y, where the y now arises from the products of elements of the last column with a, moved over to the RHS. So we now have this set of 3 equations in 3 unknowns Bx'=y,



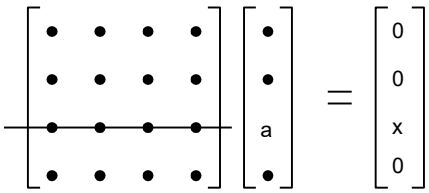
If detB ≠ 0 for the new matrix, then we can solve for the other 3 x_i components and we have found a solution. However, even if our matrix A were *known* to be rank r = N-1, we don't know anything about the *particular* 3x3 matrix determinant of B above, it may or may not be 0. If det = 0, then our approach has not yielded a solution to Bx'=y, so we don't have a solution overall to our original problem Ax = 0.

However, we could have perhaps select a different component as our forced value, such as



And this would lead us to a *different* 3x3 Bx=y matrix problem. (In this new problem, the elements of B come from columns 1,2 and 4 instead of 1,2 and 3. Again, maybe detB= 0 or maybe not. So this takes us to 4 of the 3x3 sub-determinants.

In addition to trying different positions for our forced variable, we could have started out by crossing out a different row.



So we now have 4 x 4 = 16 different ways to try to solve our equation. For each attempt, we have crossed out one row, and we have also in effect crossed out one column by putting a in some position. But this exhausts ALL possible 3x3 sub-determinants. If rank A = N-1 = 3, then at least ONE of these must be non-zero, and that one then gives us a solution.

This shows that: if rank(A) = N-1, then the equation Ax=0 has a continuum of solutions where we can set an arbitrary value to one of the components x_i (though which one, we don't know offhand) and solve for the other components. We know mechanically how we could find a vector **x** that would be such a solution, and then of course any $\alpha\mathbf{x}$ is also a solution of Ax = 0.

We can interpret this situation as saying that the dimension of the nullspace of A is 1, and in this space our solution vector **x** is our one and only basis vector. But we have not talked about nullspace so far in this document.

Define: **Nullspace of A**: the set of all vectors x that solve Ax=0 span a space of some dimension k ≤ N that is called the nullspace of A. Stakgold shows that this space is in fact a subspace of E^N. At this time, we have no theorems about nullspaces.

****

And at this point, we now *rejoin* the discussion of M&M at the bottom half of page 313.

For a given matrix A, we are always interested in these two problems: Ax=0 and Ax=y which are connected together somewhat, the first being a special case of the second. There are several cases to think about:

(a) If A is non-singular, then detA ≠ 0. We can then solve Ax=y by Cramer's Rule.

Phil Theorem 9.8 (Cramer's Rule) : If Ax=y and detA ≠ 0, then get $x_q = |\, A(q \text{ col} \leftarrow y)\,| \,/\, |\, A\,|$.

Proof: Solution is $\mathbf{x} = A^{-1}\, y = (1/\text{detA})\, \text{cof}(A)^T \mathbf{y}$. Write out the q^{th} component:

$$x_q = (1/\text{detA})\, [\, \text{cof}(A)^T]_{qp}\, y_p = (1/\text{detA})\, \Sigma_p \text{cof}(A)_{pq}\, y_p$$

But recall our earlier formula that $\det(A) = \Sigma_p\, \text{cof}(A_{pq})\, A_{pq} = \Sigma_p\, \text{cof}(A_{pq})\, [\mathbf{C}_q]_p$ where $\mathbf{C}_q$ is the q^{th} column the matrix A. If we replace this q^{th} column with the numbers y_p then we get $\Sigma_p \text{cof}(A)_{pq}\, y_p$. Let's call this matrix A(q col ← y). Then we have this Cramer's Rule result:

$$x_q = |\, A(q \text{ col} \leftarrow y)\,| \,/\, |\, A\,| \qquad q = 1,2...N$$

(b) In this same situation, the only solution of Ax=0 is x=0 (because $x = A^{-1}0 = 0$).

(c) What about Ax = 0 when detA = 0. **This is a much more complicated case**.

We know from earlier that $A\,(\text{cofA})^T = \text{detA}$, and here detA = 0, and we are trying to solve Ax = 0, so a logical choice would be to try x = the various *columns* of the matrix $(\text{cofA})^T$. If we assume (as M&M do) that at least some cofactor is non-vanishing, which would mean rank = N-1, then we get a non-zero solution by picking that column. Suppose we pick column j. We then get $x_k = (\text{cofA})^T_{kj} = \text{cofA}_{jk}$ where k is a row index as we go down column j. Any multiple $x_k = \alpha\, \text{cofA}_{jk}$ is also a solution.

Now there could be several non-zero columns of the matrix $(\text{cofA})^T$, and from each of these we could make another set of solutions in this manner. If these solutions were independent, then we would have nullity > 1, However, I think it is true that rank + nullity = N, so if rank = N-1, then nullity = 1, and so in fact there must be only one independent solution x of our equation, and we have found it. (This must be any solution one would find by the Exercise above with the 4x4 matrix.)

M&M go on to claim that we can find the ratios of all the unknowns in the solution. But if our solution above has, say, only one nonzero x component, then I guess the ratios for the other components are all zero. I don't think they are adding anything new with their comment, because $x_i = c\, \text{cof}(a_{j\,i})$ IS the solution, and it determines the ratios. I will let this confusion pass for now. Are they hinting that there is always a solution where all components are nonzero? I don't think so.

Suppose rank = N-2. In that case, all columns of $(\text{cofA})^T$ vanish, so we cannot get a candidate solution in the above manner. I suspect that in this case what you do is work with an N-1 x N-1 portion of the matrix A which portion has rank N-2. That is, you have to find a portion that has at least one cofactor

non-zero in the reduced dimensional sense. Then when you find such a portion, you apply the same argument above to find a solution x. Probably we rearrange the matrix A at the start so the first solution had basis vector 100... and our portion is then the lower N-1 x N-1 diagonal. This will make any solution found here orthogonal to the first one, and then you have found 2 solutions and nullity = 2. So I can vaguely see how this would work for any rank, and I imagine someone will have this written up for me somewhere.

****

10.10 Linear Transformations.

This entire section simply says that you can think of y = Ax in two ways. You can think of x' = Ax as giving you the coordinates of x in a new rotated coordinate system. Or you can think of x' as a new vector in the same coordinate system that has been rotated the other way. In any event, he refers to A as a linear transformation. Well, we have had this all along, but we are here thinking in a new way. Before, we regarded as Ax=y just as a matrix equation we want to solve for x. Here we are interpreting Ax = y in a certain way. We are interpreting A as an operator in a space which acts on a vector to give a new vector. This is just the Hilbert space view of matrix operations in E^N.

****

10.11 Equivalent Matrices.

Definition: B = PAQ where both P and Q are non-singular, then A and B are **equivalent**.

(a) $B = Q^{-1}AQ$ means A and B are equivalent and this is a **similarity** (= **collineatory**) transformation.

(b) $B = Q^{T}AQ$ means A and B are equivalent and A and B are **congruence** transformation.

(d) If it happens that $Q^{T} = Q^{-1}$ and Q has real elements (Q is **real orthogonal**), then $Q^{-1}AQ$ is a **real orthogonal** transformation. Since we have $Q^{-1}AQ = Q^{T}AQ$, such a transformation is both a similarity and a congruence at the same time.

(c) $B = Q^{\dagger} AQ$ means A and B are equivalent and A and B are called **conjunctive** transformation.

(e) If it happens that $Q^{\dagger} = Q^{-1}$ (Q is **unitary**) then $Q^{-1}AQ$ is a **unitary** transformation. Since we have $Q^{-1}AQ = Q^{\dagger} AQ$, such a transformation is both a similarity and a conjunct at the same time.

****

10.12 Bilinear and Quadratic Forms

Def: $A(x,y) = y^{T}Ax$ is called bilinear form.

Def: $A(x,x) = x^{T}Ax$ is called a quadratic form.

In the latter case, for any matrix A, you can construct $B = (A+A^T)/2$, then you have that $x^TAx = x^TBx$. B is a symmetrized version of A. If it happens that A(x,x) is positive for all x, then it is called **positive definite**. If in addition it can be 0, then it is **semi-definite**.

Motivation: If you look in a mechanics book like Goldstein, you find that the Lagrangian L = T - V which is kinetic - potential energy. From the Lagrangian you at once get the equations of motion and you can start to think about the problem. If, for example, you have a system of coupled harmonic oscillators, then you will find that

$$T = (1/2)\, M_{ij}\, v_i\, v_j = v^TMv \qquad V = (1/2)\, K_{ij}\, x_i\, x_j = x^TKx$$

where M is a generalization of the mass of one particle, and K the generalization of the spring k constant of the single-particle problem. These M and K are matrices, and you see that T and V are quadratic forms in their respective variables. In solving such a problem, we know that we want to find some new coordinates which hopefully diagonalize both the M and K matrices. These are the "normal coordinates" and the problem then decouples and you get your **normal modes** each with its set of eigenfrequencies. Later in this chapter, M&M will show us exactly how you can diagonalize M to the identity matrix with a certain congruence transformation, so that T will then just be a sum of squares of the components of v. The same congruence transformation will bring K to diagonal form but the diagonal elements will not all be the same.

Also, in this physics application, you can see that if matrices M and K are not at first symmetric, we can symmetrize them to make them symmetric, and T and V will have exactly the same form. Thus, we will simply assume that M and K are symmetric.

Fact: We can write the bilinear form as $A(x,y) = y^TAx = (y, Ax)$ where (a,b) is the scalar product of our Hilbert Space. If x,y and A are all real, then you can think of this as (y,Ax) using the scalar product in E^{Nr}. In this case, it is trivial to show that $(y, Ax) = (A^Ty, x)$. If in addition A is a symmetric matrix, then (y, Ax) = (Ay, x). Now, M&M *could* have defined the bilateral form as $A(x,y) = y^\dagger Ax$. In this case, we can identify this with (y, Ax) in E^c. In this case we find that $(y, Ax) = (A^\dagger y, x)$. In the case that A is a Hermitian matrix, we then have (y, Ax) = (Ay, x).

Notice again that we have three different notations to think about:
(1) In M&M, the matrix $A^\dagger$ is called the **associate matrix** of A and is written $A^\dagger$.
(2) In Quantum Mechanics, $A^\dagger$ is called the **Hermitian conjugate** of A and is written $A^\dagger$.
(3) In Stakgold, $A^\dagger$ is called the **adjoint** of A and is written A^* .

****

10.13. Similarity Transformations.

Here they show that you can interpret a similarity transformation as follows: If you map all your vectors into a new primed basis using x' = Qx, then in that new primed basis, the primed operator appears as the similarity transform of the unprimed operator, $A' = Q^{-1}AQ$,

$Ay = x$
$Q^{-1}AQQ^{-1}y \;\; Q^{-1}x$
$[Q^{-1}AQ]\, y' = x'$
$A'\, y' = x'$

****

10.14 The Characteristic Equation of a Matrix

Def: $K(\lambda) = + \det(\lambda E - A)$ is the **characteristic equation** of A. // notice choice of sign inside det
In other books, it is called the **secular equation**.

Fact: $K(\lambda)$ is a *polynomial* of degree N or less because the diagonal elements' contribution to the determinant could be of this order, and all other terms are of lesser order. If you factor this polynomial, you get the **characteristic (latent) roots**. (I would call these **eigenvalues**).

Lemma 14A: For a diagonal matrix, $\det(D) = \Pi(D_{ii})$.

Proof: Only one term of the usual n! terms survives in a diagonal matrix, and this is it!

Corollary: $\det(E) = 1$.

Lemma 14B: $\det(Q) = 1/\det(Q^{-1})$. Proof: $1 = \det(E) = \det(QQ^{-1}) = \det(Q)\det(Q^{-1})$

Lemma 14C: If $B = Q^{-1}AQ$, then $\det(B) = \det(A)$.

Proof: $\det(B) = \det(Q^{-1}AQ) = \det(Q^{-1})\det(A)\det(Q) = \det(A)$ by the above lemma.

Theorem 14D: If A and B are related by a similarity, the two matrices have the same roots.

Proof: $(\lambda E - B) = (\lambda E - Q^{-1}AQ) = Q^{-1}(\lambda E - A)Q$. By Lemma 14C, $(\lambda E - B)$ and $(\lambda E - A)$ must have the same determinant. Thus, $K_A(\lambda) = K_B(\lambda)$ so both A and B have the same eigenvalues.

Theorem 14E: If A and B are related by a similarity, the two matrix have the same trace.

Proof: $\mathrm{tr}(B) = \mathrm{tr}(Q^{-1}AQ) = \mathrm{tr}(QQ^{-1}A) = \mathrm{tr}(A)$. Second step by earlier theorem $\mathrm{tr}(AB)=\mathrm{tr}(BA)$.

****

10.15 Reduction of a Matrix to Diagonal Form

Consider $Ax = \lambda x$ which is really $(\lambda E - A)x = 0$. If $(\lambda E - A)$ is non-singular, only solution is x=0.
If $(\lambda E - A)$ is singular, then there can be non-trivial solutions x. But $(\lambda E - A)$ is only singular if

$\det(\lambda E - A) = 0$. But this is $K(\lambda)$, so $(\lambda E - A)$ is only singular if λ is root of $K(\lambda)$, which is to say, only if λ is an eigenvalue. Therefore, $Ax = \lambda x$ only has non-zero solutions if λ is an eigenvalue. The corresponding solutions are called eigenvectors, and we have $A\mathbf{x}_\lambda = \lambda\mathbf{x}_\lambda$ where λ = any eigenvalue.

If A is diagonal, then $Ax = \lambda x$ is really N separate equations of the form $D_{ii}\, x_i = \lambda\, x_i$. In this case, we can see that the eigenvalues are the D_{ii} . We have just proved this theorem:

Theorem 15A: The eigenvalues of a diagonal matrix are the diagonal values.

Theorem 15B: *If* we can find a similarity such that $Q^{-1}AQ = \Lambda$ where Λ is a diagonal matrix, then the eigenvalues of A are the diagonal elements of Λ.

Proof: We have already shown that similarity preserves eigenvalues (Theorem 14A). And we have shown that the eigenvalues of Λ are the diagonal elements (Theorem15A). QED.

Now comes the big question: <u>how to you construct a matrix Q that does the job!</u>

(a) <u>All eigenvalues are different</u>. First, we can write the eigenvalue problem $Ax_i = \lambda_i x_i$ in matrix form by constructing X with the columns x_i. Then we have $AX = X\Lambda$ (notice that the RHS is *not* ΛX). Here the matrix Λ is $\mathrm{diag}(\lambda_i)$ with the λ_i in some definite order. In Theorem 15C below, we show that if all eigenvalues are different, the eigenvectors of $A\mathbf{x} = \lambda\mathbf{x}$ are linearly independent. This tells us by another theorem above that $\det X \neq 0$ so X^{-1} exists. We can then write $X^{-1}AX = \Lambda$ and we have this brought A into $A' = \Lambda$ which is of diagonal form by a similarity X. In this transformed problem, we can say $A'\, I = I\, \Lambda$ where $A' = \Lambda$. This says that the eigenvalues are the diagonal elements of this new A', and the eigenvectors are just the usual unit vectors like (1,0,0...), since these are the columns of I. If the x_i are normalized, then X is the unique solution of $AX = X\Lambda$. If it happens that A is symmetric, it is easy to show that similarity X is real orthogonal.

The following theorem shows that the matrix X will have $\det(X) \neq 0$ and that X^{-1} therefore exists.

Theorem 15C: In the case that all the eigenvalues are different, the eigenvectors are linearly independent.

Proof: This is not instantly obvious, and M&M do not give a proof. Here is a proof given by Stakgold on his page 159. If the vectors are linearly dependent, there must be some α_i such that $\Sigma_{i=1}^{N}\, \alpha_i x_i = 0$. If you apply A to both sides, you find that this must also then be true: $\Sigma_{i=1}^{N}\, \lambda_i \alpha_i x_i = 0$. If we multiply the first equation by λ_N and then subtract that from the second equation, we get $\Sigma_{i=1}^{N-1}\, (\lambda_i - \lambda_N)\alpha_i x_i = 0$, and we have cancelled off the N^{th} term in the sum.

Now comes an induction proof. For N = 1, we know that the one eigenvector is independent (hard not to be!). Now let's *assume* that the eigenvectors are independent for N-1. Then consider the last equation written in the last paragraph. Since the $(\lambda_i - \lambda_N)$ are not 0, and since the N-1 x_i are independent, we must conclude that $\alpha_i = 0$ for i = 1 to N-1. But then our first equation in the last paragraph says that $\alpha_N x_N = 0$. Since x_N is not null, we conclude that $\alpha_N = 0$. Thus, if $\Sigma_{i=1}^{N}\, \alpha_i x_i = 0$, we find that all the α_i must be 0, and this shows that all N of the x_i are linearly independent. In our induction proof, N is the order of the matrix, so we start with N=1 and the result is then proven for any N.

Here is a simpler proof for lower cases. Suppose $x_1 = \alpha\, x_2$. Then $\lambda_1 x_1 = \alpha\lambda_2 x_2$. If you solve these two equations for x_1 and compare, you find that it is a contradiction unless the two λ are equal. Since they are not equal, you conclude that $\alpha = 0$. Thus, you could not write x_1 as a lincom of the others. This is the idea of the above proof.

The Pivot and Twist Process

(b) <u>All eigenvalues are NOT different.</u> (p 320) Here, alas, M&M have dropped the ball, so I will have to roll my own information here. Maybe I can rescue them with these words:

<u>Preliminary Comment</u>: suppose you have a vector (3,4,7) in E^3 . You *know* there is some rotation that will rotate this into a vector of the form (a,0,0), perhaps some Euler angles. We can even do it in a single rotation if we want, and there is some single rotation matrix that will do it. And |a| is the magnitude of the vector that does not change, so in this case $a^2 = 3^2 + 4^2 + 7^2$. After doing this rotation, you may want then to scale the vector to get it some desired length.

STEP 1. We know that, for any matrix A, we can certainly pick an eigenvalue λ_1 and there will be at least one eigenvector that goes with it, call it **x**. If there are several, pick one of them. Then normalize that eigenvector. As just discussed, there is *some* rotation R^{-1} (that is, some real orthogonal similarity acting on A) in E^N that will take this vector to the form $\mathbf{x'} = \{1,0,0,0..\}$ (column vector), as in our comment above. Write this as $\mathbf{x'} = R^{-1}\mathbf{x}$ where $\mathbf{x'}$ is this unit vector (we use R^{-1} to match M&M notation). Then we have:

$$A\mathbf{x} = \lambda_1\mathbf{x} \quad \Rightarrow \quad R^{-1}ARR^{-1}\mathbf{x} = \lambda_1 R^{-1}\mathbf{x} \;\Rightarrow\; (R^{-1}AR)x' = \lambda_1 x' \quad \Rightarrow \quad A'\mathbf{x'} = \lambda_1\mathbf{x'}$$

So here we have used a similarity to get from A to A', which is just a "change of basis". Now, when you apply any matrix like A' to a unit column vector (1,0...), you pick out the first column of that matrix. Therefore, according to the rightmost equation above, the first column of the matrix A' must be $(\lambda_1, 0, 0, ...)$. This is then the starting point of the M&M discussion, the picture shown bottom of page 320.

<u>Comment</u>: Nowhere do we claim that the matrix R is comprised of "the eigenvectors of A". For all we know, A maybe have a single $\lambda = \lambda_1$ and it may lack a full set of eigenvectors. We don't care! Second , notice that R is a general rotation, it probably keeps no axes fixed at all.

Summary of step 1:
(1) start with some unknown eigenvector **x**, rotate it into the unit vector $\mathbf{x'} = \{1,0,0...\}$ with some R. We normalized our eigenvector before doing this rotation.
(2) process the equation $A\mathbf{x} = \lambda_1\mathbf{x}$ into the equation $A'\mathbf{x'} = \lambda_1\mathbf{x'}$ using this R.
(3) show that the first column of A' must be $\{\lambda_1,0,0....\}$,

********** Now we pause to inject two theorems **********

Theorem 15D: If you have an eigenvalue problem $A\mathbf{x_i} = \lambda_i\mathbf{x_i}$, if you create a matrix X whose columns are the x_i, then you can write $AX = X\Lambda$ where Λ = diag $(\lambda_1, \lambda_2.....)$.

Proof: You have to draw a picture of the product XΛ and convince yourself that the columns of this product matrix are $\lambda_i \mathbf{x_i}$. Meanwhile, if you draw a picture of AX, you see that its columns are $A\mathbf{x_i}$. But we know that $A\mathbf{x_i} = \lambda_i \mathbf{x_i}$, which shows that XΛ and AX have the same columns, so they must be the same matrix. QED. Warning: notice that AX ≠ ΛX in general!!!!

Theorem 15E: Let X be the matrix of eigenvectors of the problem $A\mathbf{x_i} = \lambda_i \mathbf{x_i}$. If detX≠0, then we can diagonalize the matrix A by a similarity with the matrix X: $X^{-1}AX = \Lambda$ where Λ = diag (λ_1, λ_2.....).

Fact: When the λ_i are all different, we know from theorem 15C that detX≠ 0 so then $X^{-1}AX = \Lambda$.
Fact: When some of the λ_i are the same, I don't know the conclusion at this point.

Proof: From 15D we know that AX = XΛ. If detX≠0, then X^{-1} exists and then $X^{-1}AX = \Lambda$.

****

Resuming on page 320 bottom. I accept the form shown there. I got there with a rotation R

STEP 2. We now have our matrix A' in the following form:

$$\begin{bmatrix} \lambda 1 & b \\ 0 & B \end{bmatrix}$$

Here are the key arguments: Consider now the N-1 x N-1 matrix B. For sure, it has an eigenvalue and at least one eigenvector to go with that eigenvalue. Notice that all eigenvalues of B are also eigenvalues of A because det(A-λI) = (λ_1-λ) det(B-λI). We would like to exhausts all repeats of the eigenvalue λ_1 so they will be adjacent on the diagonal, so if λ_1 is an eigenvalue of B (that is, if A algemult(λ_1) > 1), let's pick that one for our next iteration. Now a second key point. It may in fact happen that matrix A has only *one* eigenvector for λ_1 (ie, it maybe be that geomult(λ_1) = 1 for A), and we "used it up" in step 1. Be this as it may, we still know that the smaller matrix B must have an eigenvector for λ_1 because each eigenmanifold of a matrix like B must have at least one eigenvector for each eigenvalue (we showed this elsewhere).

So for STEP 2 we just repeat what we did in STEP 1, but for the N-1 x N-1 matrix B. We apply a new rotation matrix diag(1,R1) which "clears out" the first column of B.

$$\begin{bmatrix} \lambda_1 & * & * & * \\ 0 & \lambda_2 & * & * \\ 0 & 0 & C & \\ 0 & 0 & & \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 \\ 0 & R1 \end{bmatrix}$$

This new rotation is shown on the right, and the result of similarity by that matrix on the right is the matrix on the left. Here we put λ_2 for the eigenvalue, but perhaps it was a repeat of λ_1.

So it should be very clear (it finally is to me now) that you can keep doing this process and you end up with an upper triangular (UT) matrix which has all the λ_i on the diagonal. If we select the eigenvalues in the order suggested above to exhaust each eigenmanifold of A, then we will have them grouped in that manner on the diagonal. All matrix elements above the diagonal are *, that is, undetermined. You then concatenate all your rotations together, and you have then found a rotation which takes ANY matrix to diagonal form. So we have therefore proven this theorem:

Theorem: For any real square matrix A, we can find a real orthogonal matrix R (a rotation) such that $R^{-1}AR = U$ where the matrix U is upper triangular with the eigenvalues on its diagonal.

Notice that we have NOT shown that the matrix can be put into "block diagonal form" with each block being a triangular eigenmanifold. We have just shown that the entire matrix can be put into triangular form.

Now, suppose A is a complex matrix? Consider a normalized eigenvector (a,b,c...). Can we find a unitary matrix R that "rotates" this into (1,0,0..) ? Consider $R\mathbf{v} = \mathbf{u_1}$ where R and v can have complex elements. Then this can be regarded as a set of 2N real equations in $2N^2$ real unknowns (there are N^2 matrix elements in matrix R, each can be complex). However, if we also insist that $R^{\dagger}R = 1$ (R is unitary), then that is another set of 2N real equations in the same set of $2N^2$ unknowns, so overall, we have 4N real equations in $2N^2$ unknowns. If N = 2, we have 8 equations in 8 unknowns, so there is only one solution. For N=3 we have 12 equations in 18 unknowns, so there should be many solutions. Saying that R is unitary says that |detR| = 1. I have not worked out all the fine details here, but it seems very likely to me that yes, you can always find at least one unitary matrix which "rotates" any complex normalized vector v into the real unit vector u_1.

Assuming this is in fact true, we can apply the same argument as above, with "real orthogonal" R replaced by "unitary R". At each stage, just as before, we clear out the first column section. So I think this then proves the following:

Theorem (Schur's Theorem) : For any square matrix A, we can find a unitary matrix R such that $R^{-1}AR = U$ where the matrix U is upper triangular with the eigenvalues on its diagonal. This appears in Allen's lecture notes in Section 4.2. as Theorem 4.2.1, and he gives a very simple induction proof.

Now, M&M go on to claim a more stringent fact: that any matrix A can be brought to block diagonal form where the blocks are the eigenmanifolds of each λ_i , and where each block is triangular. They claim it, but do not prove it except with some weak arm-waving. Allen makes this same claim in his Theorem 4.2.2. He runs into the same problems that I did (see my separate block diag notes) regarding how you "clear out" the upper right regions, so his proof is a little messy like mine. Of course from Matthews, I now know that we can write the matrix A as a direct sum of the generalized eigenspaces $\mathrm{Ker}(p^b)$, and that at once means you have the diagonal block form, ie, there is a basis where A is transformed into block form. Then in each of those blocks you can do the Schur Theorem to get each block into triangular form. And I also know that we can go further still and reduce each of these triangular blocks into a sum of elementary Jordan blocks, a subject not mentioned at all by M&M. So I will just state the theorem here:

Theorem (Block Triangular Form) : For any square matrix A, we can find a unitary matrix R such that

$R^{-1}AR = B$ where $B = \text{diag}(U_1, U_2 ....)$ where we have one block U_i per eigenmanifold, and where each of these blocks is upper triangular, having the eigenvalue of the manifold in the diagonal.

When each eigenvalue occurs only once, we know that *a* transformation R is equal to the matrix you get by assembling the eigenvectors of $Ax = \lambda x$. When there is degeneracy, we know it is certainly possible to *not have* N linearly independent column vector eigenvectors, so in such a case, R still exists, but it cannot possibly be the same matrix you get by assembling eigenvectors. [It consists in fact of all the generalized eigenvectors.] We can still associate the non-degenerate columns of R with the singlet eigenvalues (they have simple column unit vectors matching the location of λ in the matrix), but the columns which go through a degeneracy box are hard to interpret. I think all the actions getting to the overall triangular form really are just rotations regardless of degeneracy. (yes)

If the matrix A is **symmetric (Hermitian)**, it will be shown two sections ahead that the block-diagonal form itself can become *completely* diagonal if the correct similarity is used, even if there is eigenvalue degeneracy. The claim is that our R above is not unique (see note above, it is very true!). The argument is that one knows there is a congruence transform which fully diagonalizes any symmetric A, and you write $Q^T AQ = \Lambda$. It is claimed that a special modified version of Q can be generated by doing G-S on Q, and we can call this R so $R^T AR = \Lambda$. Due to the G-S procedure, this R has the property that $R^T R = 1$, so R is now real orthogonal. But in this section we showed that $R^{-1}AR$ = block diagonal form. But in the future section they show $R^T AR = \Lambda$. Thus, the off diagonal elements of the block diagonal form must vanish when A is symmetric and we use this R as our similarity. The point is that even if you have λ degeneracy, you can diagonalize A completely using R (if symmetric), and eigenvalues are preserved to boot. I guess what we have to leave behind is the idea that the columns of this R are the eigenvectors.

****

(page break)

Just for the record, here were the two main problems I was having decoding the M&M text:
(1) why can you clear F out? How do you do it?
(2) as you continue with a new eigenvalue λ_2, why do you get another isolated "box" instead of a continuation of the matrix A_1 to something ever larger (but with λ_2's appearing). I have the feeling that M&M are "faking" this derivation and there is something very important that they have not said.

****

10.16 Congruent Transformations.

This refers to $B = Q^T A Q$.

In the previous section, we were focused on the eigenvalue problem and degeneracy and how to simplify the matrix A which appears in $Ax = \lambda x$. We accomplished the block diagonal form by doing various similarity transformations of the form $A' = X^{-1}AX$.

In this section, we ignore the eigenvalue problem, and work instead with a *different* prototype problem, and this is the one of trying to simplify a quadratic form. As noted earlier, in such a form your matrix of interest is already sandwiched between two x vectors, it is not free-standing. This form does not change if you symmetrize the sandwiched matrix, so you can always think of A as symmetric. It turns out that we will simplify the quadratic form with a transform of the form $B = Q^T A Q$, where $Q^T \neq Q^{-1}$, so we are not using any similarity transforms in this section of the chapter. In fact, Q^{-1} is never even mentioned.

So the key starting point is to realize that a quadratic form x^TAx is unchanged if you replace the matrix A with the matrix $B = Q^T A Q$. If we can get B to be diagonal Λ, then the quadratic form has no cross terms ($\Sigma B_{ii} x_i^2$) and becomes much easier to work with when solving Lagrange equations etc.

So they first show that if you change variables by $x = Qy$, you can compensate by doing $B = Q^T A Q$ and then $x^TAx = y^TBy$.

So, what we are really discussing here (couched in the bilinear form example) is HOW you diagonalize a symmetric matrix A. This is done in the form $\Lambda = Q^TAQ$ where Q turns out to be upper triangular with 1's on the diagonal. [Note inserted: of course since A is symmetric = Hermitian, we know from elsewhere that we can diagonalize it with a real orthogonal Q with $Q^T = Q^{-1}$. But in this chapter we are going to first see there are a large class of non-RO congruences Q that do the job $\Lambda = Q^TAQ$.]

A Process of finding a Q to diagonalize symmetric A. (p 322) **Pivot and Clear**

You see how A is written at the start. You see the specific matrix Q_1 which *isolates* the upper-left element A_{11} in the matrix A. By isolate, I mean puts a pair of zeros as shown. I did this on scratch paper and agree exactly with the results shown. This matrix Q_1 is shown in more detail on page 323.

You now repeat this operation in the N-1 size submatrix of A". You operation is $Q_2 = \text{diag}(1,q_2)$, and this causes two things: (1) In the new matrix, another 0 will occur in the first row and column, just due to the fact that you have $\text{diag}(1,q_2)$, so you don't "mess up" your diagonalization of the A_{11} element. (2) Within the N-1 space, we will isolate element A_{22} on the diagonal with a zero on each side.

So it is now very clear to me that as you keep applying these Q's, you will eventually diagonalize A after N steps. And I agree that the final Q will have the shape shown with 1's on the diagonal.

Note that this procedure is VERY different from our Pivot and Twist where we applied mysterious rotations to bring a partial column of an unseen eigenvector x to the (***100000) form. Here the transformations Q_i are completely explicit. Also, note that $Q^T AQ$ does not preserve eigenvalues.

Here, we are doing a sort of Pivot and Clear operation. Authors point out that instead of starting with the first row, somehow we can start with any other row with $A_{ii} \neq 0$ with a different look to the Q's. Their point is that there are very many different matrices Q which do the job.

What happens if $A_{11} = 0$? We are referred to a footnote, so the fix is non-trivial!

Note: in my little derivation, it is *crucial* that the lower left part of A be V^T. Otherwise, we fail to clear out the lower left part of A by doing Q_1. Thus, it is crucial that A be symmetric. So far then we have some results to report out:

Lemma: The product of two upper triangular matrices is upper triangular.
Proof: I just looked at it on paper and it seems pretty clear. This helps show why Q is this way.

Ignoring the technical problems like $A_{11} = 0$, we have now proven this theorem:

Theorem 16A: A symmetric matrix A can be diagonalized by a congruence transformation $B = Q^T A Q$ where B is diagonal and where Q is an upper triangular matrix.

Fact: This is a useful fact if you are trying to compute the quadratic form $x^T Ax$, because the result is then simply $\Sigma B_{ii} y_i^2$ where $x = Qy$.

Fact: A congruence transformation does NOT preserve eigenvalues because $\det(Q^T Q) \neq 1$ in general. But in this section, we don't care about the eigenvalue problem!

Fact: As they note, the Q and B_{ii} you obtain are not unique! The *value* of the quadratic form however will be unique since it is always going to be equal to $x^T Ax$ regardless of Q. So the terms in the sum change, but the sum stays the same.

On page 323 top you see the diagonalized form for something like kinetic energy where there are no cross terms between the ξ variables. The diagonal elements are $D_{ij} = \alpha_j$ for short. We have transformed to a new "coordinate system" with ξ variables where $x = Q\xi$. This is the "principle axis" change. In the new "normal coordinates" the problem simplifies. Later on this page he shows how you can compute the α_j directly from sub-determinants called **discriminants**, a sort of Scheid-like topic. Then finally he notes that you can do a final renormalization of your normal coordinates to get rid of the α's, no big deal.

Note: He does say that Q is not unique as a solution to achieve diagonalization, but the particular Q he shows which is upper triangular *is* in fact unique because we constructed it by an exact specification from the matrix A. With this particular Q, we diagonalize to a particular D.

There are many other Q's that do the job, and the resulting D's are all different!!! This is different from the Pivot and Twist where we were always preserving eigenvalues. Even thought D's might be different, the quadratic form $x^T Ax$ will have the same value when written as $y^T Dy = \Sigma D_{ii} y_i^2$.

Review: There are many congruences Q that bring a symmetric A to a diagonal form D. This diagonal form of D is nice because then $x^T Ax = y^T Dy = \Sigma D_{ii} y_i^2$ or even $\Sigma \eta_i^2$. These y or η are normal

coordinates. The Q and D are non-unique, but the value of $x^T Ax$ is the same for any Q and D. This chapter shows one method for constructing a Q that does the job.

****

10.17 Orthogonal Transformations

Now we are closer to home. We open with a few facts like

Theorem 17A: if R is real orthogonal, meaning $R^{-1} = R^T$, then det(R) = ±1. Discussion of inversions.
Proof: $1 = \det(E) = \det(R^T R) = \det(R^T)\det(R) = \det(R)^2$ QED.

Theorem 17B: R preserves the length of a vector, so that $\| y \|^2 = y^T y = \| x \|^2$.
Proof: very obvious.

On page 325 we then have the slightly obscure discussion which is basically this: There are many congruences Q which bring A to some diagonal form. Of these many, there is one congruence which has these properties: (1) this particular Q is real orthogonal; (2) the resulting diagonal matrix is Λ. It is suggested that one way you could get to this Q could be to first find Q such that $Q^T AQ = I$, then you find the particular Gram-Schmidt orthonormalization process which (a) converts Q to Q' which is orthogonal and (b) which gives the exact Q' which diagonalizes A. Well, OK, but not very practical.

We can look at it the other way as well. In our similarity section doing Pivot and Twist, we had lots of freedom in deciding the rotations which make up W. In the case that A is symmetric, we know that the successful diagonalizing matrix W will be the eigenvectors of A in any normalization we want. There is one choice of eigenvectors (another GS) which makes X be not just a similarity, but also a congruence. so out of many similarities, one is also a congruence. But in this case, for this one special similarity that is also a congruence, conservation of symmetry tells us that the Pivot & Twist resulting matrix must be diagonal. In the previous paragraph we said that out of many congruences, one is also a similarity. All a bit obscure

OK, now we are on page 326, and once again, they are back to their favorite subject of quadratic forms. This time we are going to worry about TWO quadratic forms at the same time. We first reduce A to D using a Q. This same Q is not going to the other form B diagonal but instead converts it to some C. OK fine. We then renormalize our y components as shown so **y** is now replaced by **η**. This puts the D form into simplest sum of squares form as shown. C becomes C' in the η basis. At this point we find R such that R will diagonalize C', and this takes us from basis η to ξ . In the final basis, we find that D is just the sum of component squares, while C is similar, but with the λ_i of the Λ that it became.

All this relates to problems in molecular physics where you are looking for normal modes of vibration, but I have forgotten that part of my past completely, so my motivation is weak.

Next comes a sudden new topic. What happens if we try to diagonalize an orthogonal matrix R. What do I know about the eigenvalues of such an R? Not much. Have to study det(λE-R) = 0. Once again, M&M send me off to the woods to fetch this theorem:

Theorem 17C: The eigenvalues of a real orthogonal (or Hermitian) matrix all have $|\lambda| = 1$. Complex eigenvalues must occur in pairs.

Proof: Write $Rv = \lambda v$. Take conjugate of this to get another equation $v^\dagger R^\dagger = v^\dagger \lambda^*$. Then multiply these two equations together to get $v^\dagger R^\dagger Rv = |\lambda|^2 v^\dagger v$. If R is real orthogonal, then $R^\dagger = R^T = R^{-1}$ and we have then shown that $v^\dagger v = |\lambda|^2 v^\dagger v$, QED statement 1.

Now suppose $Rv = \lambda v$. Then $R^*v^* = \lambda^* v^*$. But R is real, so $Rv^* = \lambda^* v^*$. Thus, if λ is an eigenvalue with vector v, then λ^* is also an eigenvalue and its vector is v^*. QED statement 2.

We saw a nice example of this in Stakgold page 158 example 2. The eigenvalues of the usual 2x2 rotation matrix are $\exp(\pm i\theta)$.

Even knowing this theorem, I am at a complete loss regarding (10-43). First of all, how do I know that you can even diagonalize a matrix R that is real orthogonal? This section is revealing two theorems that I have not yet investigated:

(1) If you hunt for the eigenvalues of a general rotation matrix, they are all of the form $e^{i\phi}$ as noted above, but in certain cases you know that this is +1 or -1. The example in Stakgold is a 2x2 simple rotation and it falls into the first of the four classes shown. It is not at once obvious to me how you show these + 1 and -1 special values exist.

(2) The fact that you can diagonalize any real orthogonal R by means of a unitary matrix U is new to me. This is not something I have tried to do before.

Regarding #2, consider the eigenvalue problem $RX = X\Lambda$. In this case, we anticipate that the elements of X will be complex, since the eigenvalues are generally complex. If we then orthonormalize the columns of X "as usual", it becomes unitary, the generalization of real orthogonal. OK, so this is then why you can diagonalize a real orthogonal matrix with a unitary matrix, fine! It is proved.

Theorem 17D: Any real orthogonal matrix R can be diagonalized to Λ by a unitary matrix U. The eigenvalues are complex but unimodular.

****

10.18 Hermitian Vector Space

Nothing new to me here except the way they write the scalar product: $(x,y) = x^\dagger y$

****

10.19 Hermitian Matrices.

Theorem 19A: (x,Ax) is real if A is Hermitian

Theorem 19B: The eigenvalues of A are real if A is Hermitian

Theorem 19C: A Hermitian matrix can be diagonalized by a unitary similarity.

Theorem 19D: Hermiticity is conserved under a similarity by any unitary matrix.
Just the way symmetry is conserved under similarity by a real orthogonal.

Theorem 19E: If two matrices commute, they can both be diagonalized by the *same* unitary similarity. The converse is also true.

This is of course basic to QM and conserved quantum numbers. M&M prove only the converse, that if you can diagonalize both at once, then they must commute, and we have a footnote for the other way.

****

10.20 Unitary Matrices.

All the usual stuff, including this (with proofs exactly as given earlier).

Theorem 20A: Any unitary U can be diagonalized to Λ by a unitary matrix V. The eigenvalues are complex but unimodular.

****

10.21 Summary on Diagonalization.

Here are the rules:

(1) a Hermitian (sym) matrix can be diagonalized by a unitary (r.o.) similarity

(2) A Unitary (r.o.) matrix can be diagonalized by a unitary similarity

(3) A matrix with all-different eigenvalues can be diagonalized. In this case, recall that the eigenvectors are orthogonal (Theorem 8 in Some Matrix Theorems). If you normalize them, it seems to me that you get a X to be unitary, so I would say in this case you can diagonalize by a unitary similarity as in the other cases. But they don't say that for some reason, their last words of this long chapter!

****