Some Matrix Theorems PhL 12.11.04

This is all meat and potatoes matrix stuff, keep this document! It was written while I was reading M&M ∂chapter 10. See separate notes for details of that chapter.

A Warm-up Example

The eigenvectors of a symmetric matrix are orthogonal (Theorem later). Let's look at a simple example.

$$B = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \text{ so } B-\lambda E = \begin{pmatrix} a-\lambda & b \\ b & c-\lambda \end{pmatrix} \text{ and secular says } (a-\lambda)(c-\lambda) - b^2 = 0$$

which we rewrite as $\lambda^2 - (a+c)\lambda + (ac-b^2) = 0$, therefore $\lambda_\pm = (1/2)\left[(a+c) \pm \sqrt{(a-c)^2 + 4b^2} \right]$.
Notice that we can factor this as $(\lambda-\lambda_+)(\lambda-\lambda_-) = 0$ so that tells us $\lambda_+\lambda_- = ac-b^2$ and also that
$(\lambda_+ + \lambda_-) = a+c$.

Aside: Notice that the eigenvalues are the same only if a=c and b=0, so the only possible degenerate 2x2 matrix is a multiple of the identity. This seems a surprising result to me, but it must be true. (See theorem 4S below)

The eigenvalue equations written out are these: (two pairs of equations)

$$\begin{aligned} (a-\lambda)x + by &= 0 \\ bx + (c-\lambda)y &= 0 \end{pmatrix}$$

We cannot of course to Cramer's to solve since det=0. Assuming $b \neq 0$, we can solve second one for x to get $x = [(\lambda-c)/b]\, y$. When this is put into the first equation, it becomes 0y=0, so any y works. Therefore, our eigenvectors are these:

$$\begin{pmatrix} [(\lambda_\pm - c)/b]\, y \\ y \end{pmatrix} = k \begin{pmatrix} [(\lambda_\pm - c)/b] \\ 1 \end{pmatrix} = x_\pm$$

So if we install our two eigenvalues, we get two eigenvectors and they are different. Are they orthogonal? That is our big question. The answer is YES. We find by direct calculation that $\mathbf{x}_+ \bullet \mathbf{x}_- = 0$.

********** **And now on with the show** **********

Theorem 0: A real matrix A can have complex eigenvalues and complex eigenvectors.

An example is $A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ which has Maple = [a + I b, 1, {[-I, 1]}], [a - I b, 1, {[I, 1]}]
where format is eigenvalue, multiplicity, eigenvector. You see the complex eigenvalues and the complex eigenvectors. Later we see that a symmetric real matrix A has everything real.

Definition: If Ax=0, the **nullity** of A is the dimensionality of the nullspace of A, which is in turn the number of distinct non-zero vectors that solve Ax = 0.

Definition: The **geometric multiplicity** of an eigenvalue λ of the problem $Ax=\lambda x$ is the nullity of $(A-\lambda E)$, that is, it is the dimensionality of the nullspace of $(A-\lambda E)$. It is thus the number of independent non-zero solutions x of the equation $(A-\lambda E)x = 0$.

Definition: The **algebraic multiplicity** of an eigenvalue is the number of times it appears in the factored secular equation.

Theorem 1 : Geometric multiplicity $m \leq$ Algebraic multiplicity k.

This is not quite as obvious as it looks, so we peek at Stakgold page 157 for a proof. Start with geo = m. You then have m distinct eigenvectors for $Ax = \lambda_1 x$. Stakgold tells us to "go to a basis where the matrix A has the form shown", in which form these m eigenvectors are associated with the first m unit vectors, and the matrix A has an upper left corner that is just λ_1 times E_m.

How do we know that this is possible to do? I will answer this "constructively" by showing how you would get there from an arbitrary starting point. Apply M&M's Pivot and Twist process from page 320 where you take for your first eigenvalues the m copies of λ_1, and get the matrix A into the form shown in (10-41) but where the triangular matrix is m x m with all λ_1, and we have not processed the rest of the matrix. Getting to this point corresponds to some change of basis x' = Qx from the starting basis. If we now consider just this mxm piece of the larger matrix, we can think of the local problem $AX = X\Lambda$ and we have assumed in the Pivot and Twist process m column vectors for X, these are our m eigenvectors. They are linearly independent (since m = geo) so X is invertible. Thus, we know we can diagonalize in this local region by saying $X^{-1}AX = \Lambda$. But this is just another basis change x" = Xx'. Thus, we have arrived at a basis x" where the matrix indeed has the form shown on page 137 of Stakgold. When we enlarge to diag(X,1), the B area gets altered by the diagonalization stage, which is fine, and it ends up being Stakgolds B.

Now we can finish Stakgold's proof. With A in the form shown, the secular equation has the form that $(\lambda-\lambda_1)^m \det(C-\lambda E) = 0$. Since the second factor might have more roots at λ_1, we conclude the result we need, which is $k \geq m$. Typically, in fact, there will be more roots in the $\det(C-\lambda E)$ part.

Notice that this proof did not in fact require our final diagonalization step. Once the submatrix is in triangular form, we get the same secular equation just quoted, and we draw the same conclusion.

Corollary: For any matrix A of order N, if λ is an eigenvalue of algebraic multiplicity k, then there can be at most k distinct eigenvectors associated with λ.

Theorem 2: For any NxN matrix A, if λ is an eigenvalue of algebraic multiplicity k, the number m of distinct eigenvectors associated with λ is given by $m = N - \text{rank}(A-\lambda E)$ and $1 \leq m \leq k$.

Proof: the number m is the nullity of $A-\lambda E$ and according to Theorem R6 (way below), $m = N - \text{rank}(A-\lambda E)$. Although it would seem at first that m could range from 0 to N, it really can only range from 1 to k. The upper limit comes from Theorem 1 above which says $m \leq k$. The lower limit occurs because, when λ is an eigenvalue, $\text{rank}(A-\lambda E) < N$ since $\det(A-\lambda E) = 0$, so the largest $\text{rank}(A-\lambda E)$ can be is N-1, in which case m = 1.

Corollary: There is always at least one eigenvector for a given λ. (just because $m \geq 1$) A matrix always has N eigenvalues, but they can clump in sameness groups, eg, all could be the same.

Question: Does a Hermitian NxN matrix always have N eigenvectors?

The answer is yes, and this time we go to Stakgold page 160. Again, things are very subtle.

For warm up, consider this example of algebraic multiplicity = 2, which is NOT Hermitian,

$\begin{pmatrix}\lambda_1 & c\\ 0 & \lambda_1\end{pmatrix}\begin{pmatrix}y\\ z\end{pmatrix} = \lambda_1\begin{pmatrix}y\\ z\end{pmatrix} = \begin{pmatrix}\lambda_1 y+cz\\ \lambda_1 z\end{pmatrix}$. The upper equation says z = 0 so we get $\begin{pmatrix}y\\ z\end{pmatrix} = \begin{pmatrix}y\\ 0\end{pmatrix} \sim \begin{pmatrix}1\\ 0\end{pmatrix}$.

This is the ONLY eigenvector. Now consider: $\begin{pmatrix}\lambda_1 & c\\ 0 & \lambda_1\end{pmatrix}\begin{pmatrix}0\\ 1\end{pmatrix} = \begin{pmatrix}c\\ \lambda_1\end{pmatrix} = c\begin{pmatrix}1\\ 0\end{pmatrix} + \lambda_1\begin{pmatrix}0\\ 1\end{pmatrix}$. This is an example of something important for this theorem. Eigenvalue λ_1 has a manifold of dimension 1, meaning geo m = 1, even though our alg k = 2. Call this manifold M_1 . The perp space $M_1^{\perp}$ clearly has dimension 1. But the perp space is not "closed" under the action of A. We have y = $\begin{pmatrix}0\\ 1\end{pmatrix}$ in $M_1^{\perp}$, but Ay = $c\begin{pmatrix}1\\ 0\end{pmatrix} + \lambda_2\begin{pmatrix}0\\ 1\end{pmatrix}$ which means the result has a piece in M_1 and a piece in $M_1^{\perp}$. This says that $M_1^{\perp}$ is *not* a linear manifold. If it *were*, we would consider Ay =λ_2y in this manifold, and we know that there is always one eigenvalue and one eigenvector in a manifold. But this would lead to a contradiction because there IS no other λ_2 in our secular equation. But since M_1 is not a manifold, we have no contradiction. In this example, then, you cannot "partition" the space E^2 into two eigenmanifolds of A each of dim = 1. Another way to say this is that you cannot partition E^2 into a set of separate nullspaces of (A - λE). The nullspace has dim = 1, and the other dim is not a separate nullspace of some other eigenvalue.

Now we are ready to move on.

Theorem 3: Let A be an NxN complex matrix which has only one eigenvalue λ and its corresponding single eigenmanifold. Matrix A therefore has algebraic multiplicity k = N. In general, the geometric multiplicity m is m ≤ N (another theorem). **However, if A is Hermitian, then m =k = N.**

Proof: Let M_1 be the eigenmanifold of λ_1 and assume some geometric multiplicity m = m_1 < N. Consider vector y in $M_1^{\perp}$. When A is Hermitian, Ay is also in $M_1^{\perp}$ (as we shall show below), so we consider the reduced problem Ay=λy in Hilbert space $M_1^{\perp}$ which has dimension N - m_1. This problem must have at least one eigenvalue and eigenvector. The eigenvalue cannot be λ_1 because all vectors with λ_1 are in M_1. The eigenvalue cannot be some other value $\lambda_2 \neq \lambda_1$ because A has no eigenvalue other than λ_1. Thus, the space $M_1^{\perp}$ must be null, which means m_1 = N.

The reason Ay is in $M_1^{\perp}$ if y is in $M_1^{\perp}$ is this: (x,Ay) = (Ax,y) = (x',y) = 0 for any x. We are using x and x' as being in M_1 and y in $M_1^{\perp}$. Note that A being Hermitian is used in the first step. We have seen in our example earlier that if A is not Hermitian, then Ay need not be in $M_1^{\perp}$ and generally has components in both M_1 and in $M_1^{\perp}$. Thus, in that case, $M_1^{\perp}$ is not a closed space under the action of A. That is, A does not have domain = range = some Hilbert space.

Theorem 4: The only Hermitian matrix having just one eigenvalue λ is $A = \lambda E$.

Proof: If A is Hermitian with only one λ_1, then rank$(A-\lambda_1 E) = N-m = 0$ which means $A-\lambda_1 E = 0$ so $A = \lambda_1 E$. There is another way to see this. Start with any matrix A and apply the M&M Pivot and Rotate process to get it to triangular form. The secular for this matrix is $(\lambda-\lambda_1)^N = 0$. This triangular matrix is equivalent to the starting matrix, just a basis change away. So any matrix with a single eigenvalue can be put into this form. But such a triangular matrix is only Hermitian of all off diagonal elements vanish.

Theorem 5: If A is Hermitian, then **each eigenmanifold has k = m**. That is to say, for each eigenmanifold, the geometric multiplicity has its maximal value and it equal to the algebraic multiplicity.

Proof: We just continue the general idea of Theorem 3 above, see also Stakgold page 160. We pick a first manifold and assume it has geo mult = m_1 and that is our M_1. We then look at our reduced problem in $M_1^{\perp}$ of dim N - m_1. We find the next manifold of geo = m_2 and that is M_2 so we then look at $(M_1+M_2)^{\perp}$. We continue on and in this way we find that $m_1 + m_2 + ... = N$. The key is that each of these perp spaces is closed under A due to the Hermiticity of A.

Now, we know that $k_1 + k_2 + ... = N$ because we are allowing complex roots in our analysis here, and the secular equation is a poly of degree N. But suppose we were doing a "real version" of our proof here, then we would only know that $k_1 + k_2 + .. \leq N$, the idea being that we don't count complex roots if there were some. Consider then this collection of information:

$$m_1 + m_2 + ... = N$$
$$k_1 + k_2 + .. \leq N$$
$$m_i \leq k_i$$

We then know that $N = m_1 + m_2 + ... \leq k_1 + k_2 + ... \leq N$. It would follow that $k_1 + k_2 + .. = N$ and we get this result anyway.

We can now show that $m_i = k_i$ for all i, meaning that each manifold is "geometrically maximal". Assume this is not true, and introduce some $\delta_i \geq 0$ such that $m_i = k_i - \delta_i$. Then since $\Sigma m_i = \Sigma k_i = N$, we have $\Sigma\delta_i = 0$. Since all δ_i are ≥ 0, we have that $\delta_i = 0$ for all i. Thus, it must be that $m_i = k_i$ for all i.

Theorem 5A. If A is Hermitian, then the nullspaces of $(A - \lambda E)x=0$ *partition* the whole space. You can then decompose any vector f into components each of which is in one of these nullspaces. These nullspaces are of course the eigenmanifolds of $Ax = \lambda x$. There are no vectors x in the whole space that are *not* elements of one of the nullspaces. For A non-Hermitian, we showed a 2x2 example earlier where we *could* find elements that were in non of the nullspaces.

Proof: These statements all follow from the comments in the proof of the previous theorem. This partitioning idea I think is a main feature of Hermitian A.

Theorem 6: An NxN Hermitian matrix has **N distinct eigenvectors.**

Proof: This follows from the previous theorem. Each manifold has the maximal $m_i = k_i$ and the total number of distinct eigenvectors is $\Sigma m_i = N$. Notice that a non-Hermitian matrix can have less than N eigenvectors.

Theorem 7: All eigenvalues of a Hermitian matrix **are all real**.

Proof: (guided by Stakgold page 181). $Ax = \lambda x$ is the eigenvalue problem, so

line 1: $(Ax,x)^* = (x,Ax) = \lambda(x,x)$

line 2: $(Ax,x)^* = [\,(x,Ax)\,]^* = [\,\lambda\,(x,x)\,]^* = \lambda^*(x,x)^* = \lambda^*(x,x)$

Comparison shows that $\lambda = \lambda^*$ so λ is real. From the first line, then (x,Ax) is also real.

Theorem 8: For any Hermitian matrix A in complex space, the eigenvectors corresponding to different eigenvalues are **orthogonal**.

[Comments: If A is *not* Hermitian, we don't have $(x_1,A^\dagger x_2) = (x_1,Ax_2)$ and the proof below *fails* and eigenvectors in different eigenmanifolds are generally NOT orthogonal, even in the case that all eigenvalues are different. In this case the eigenvectors at least are linearly independent, though not orthogonal.]

Proof: Write

$$Ax_1 = \lambda_1 x_1$$
$$Ax_2 = \lambda_2 x_2 .$$

Then

$$(x_2,Ax_1) = \lambda_1(x_2, x_1)$$
$$(x_1,Ax_2) = \lambda_2(x_1, x_2).$$

We now alter the first of these equations as follows:

$$(x_2,Ax_1)^* = \lambda_1^*(x_2, x_1)^* \;\Rightarrow\; (Ax_1,x_2) = \lambda_1^* (x_1, x_2) = (x_1,A^\dagger x_2)$$

If A is Hermitian then $(x_1,A^\dagger x_2) = (x_1,Ax_2)$ and our two equations are now

$$(x_1,Ax_2) = \lambda_1^* (x_1, x_2)$$
$$(x_1,Ax_2) = \lambda_2 (x_1, x_2)$$

Subtract to find that

$$[\,\lambda_1^* - \lambda_2\,](x_1, x_2) = 0.$$

From a previous theorem we know that all λ are real, so this becomes

$$[\,\lambda_1 - \lambda_2\,](x_1, x_2) = 0$$

Therefore, if two eigenvalues are different, their eigenvectors must be orthogonal.

Theorem 9: A Hermitian NxN matrix A has N eigenvectors which are **linearly independent** and thus form a basis in $E^{N.}$ It is possible to make the eigenvectors form an **orthogonal** or **orthonormal** set.

Proof: The m_i eigenvectors for each eigenmanifold are linearly independent by assumption, since m_i is the geometric multiplicity of that manifold. We know that eigenvectors from different eigenmanifolds are orthogonal to each other from Theorem 8 (below). Thus, the total collection of eigenvectors from all the manifolds is linearly independent. According to theorem 3, this collection contains N eigenvectors because each manifold is geometrically maximal, $m_i = k_i$. Thus, our set of N eigenvectors is linearly independent. If put into a matrix, that matrix would be therefore invertible.

Within each eigenmanifold, we can do a Gram-Schmidt orthogonalization process if we want to make the local vectors be either orthogonal or orthonormal. Thus, the overall set of eigenvectors can be made orthogonal or orthonormal. It is good to remember that each eigenvector has a scaling degree of freedom for the orthogonal sets.

Theorem 10: The components of the eigenvectors of a Hermitian matrix A are in general complex.

Consider $Ax = \lambda x$. Even though λ is known to be real, A can have complex elements, so when we solve the system of N equations in N unknowns, we generally get complex x.

Theorem 11: An NxN Hermitian matrix A can be diagonalized by a unitary similarity X, and X is unique only if there are no degenerate eigenmanifolds for A. A can also be diagonalized by a huge set of non-unitary X similarities.

Proof: The eigenvalue problem is $AX = X\Lambda$. Since the columns of X (the N eigenvalues x_i) are linearly independent, $\det X \neq 0$ and we can write $X^{-1}AX = \Lambda$. This is true for any choice of how we normalize those eigenvectors. If we choose an orthonormal set for the x_i , which is possible by Theorem 9, then we have $X^{\dagger}X = 1$ and X is then a unitary matrix. The reason is this:

orthonormal eigenvectors =>

$$\delta_{k,\ell} = (x_k, x_\ell) = \Sigma_i (x_k)_i^* (x_\ell)_i = \Sigma_i X_{ik}^* X_{i\ell} = \Sigma_i (X^{\dagger})_{ki} X_{i\ell} = (X^{\dagger}X)_{k\ell} \quad => 1 = X^{\dagger}X$$

There are many similarities X we can find to diagonalize A, but at least one of them will be unitary. The unitary X is not unique if there are degenerate manifolds of A. Within each, we are free to orthogonalize in a variety of ways, each giving different X. If there is no degeneracy, then there is only one unitary X that does the job.

****

We can now provide real **symmetric versions of the above theorems**. Every real symmetric matrix is Hermitian, so all the theorems pass through. Since $Ax = \lambda x$ and A and λ real, it does not make much sense to have x be complex. Making it complex just replicates the solutions in the real and imaginary parts.

Theorem 1 : Geometric multiplicity $m \leq$ Algebraic multiplicity k.

Corollary: For any matrix A of order N, if λ is an eigenvalue of algebraic multiplicity k, then there can be at most k distinct eigenvectors associated with λ.

Theorem 2: For any matrix A of order N, nullity(A) = N - rank(A)

Theorem 2A: For any NxN matrix A, if λ is an eigenvalue of algebraic multiplicity k, the number m of distinct eigenvectors associated with λ is given by $m = N - \text{rank}(A-\lambda E)$ and $1 \leq m \leq k$.

Theorem 3S: Let A be an NxN real matrix which has only one eigenvalue λ and its corresponding single eigenmanifold. Matrix A therefore has algebraic multiplicity k = N. In general, the geometric multiplicity m is $m \leq N$ (another theorem). **However, if A is real symmetric, then m =k = N.**

Theorem 4S: The only real symmetric matrix having just one eigenvalue λ is $A = \lambda E$.

Theorem 5S: If A is real symmetric, then **each eigenmanifold has k = m**. That is to say, for each eigenmanifold, the geometric multiplicity has its maximal value and it equal to the algebraic multiplicity.

Theorem 6S: An NxN real symmetric matrix has **N distinct real eigenvectors.**

Theorem 7S: All eigenvalues of a real symmetric matrix **are all real**.

This theorem is not very obvious when you just think of the secular equation. A priori, it could have complex roots. It is not obvious how you would directly prove that there are no complex roots just using the property $A_{ij} = A_{ji}$ in the equation $\det(A-\lambda I)=0$. In fact, none of these theorems is very obvious.

Theorem 8S: For any real symmetric matrix A , the eigenvectors corresponding to different eigenvalues are **orthogonal**.

Theorem 9S: A real symmetric NxN matrix A has N real eigenvectors which are **linearly independent** and thus form a basis in $E^{N.}$ It is possible to make the eigenvectors form an **orthogonal** or **orthonormal** set.

Theorem 10S: The components of the eigenvectors of a real symmetric matrix A can be taken real with no loss of generality.

Theorem 11S: An NxN real symmetric matrix A can be diagonalized by a real orthogonal similarity X, and X is unique only if there are no degenerate eigenmanifolds for A. A can also be diagonalized by a huge set of non-orthogonal X similarities.

****

Some Theorems about Rank.

Definition: The **rank** of a matrix is the size of the smallest non-zero sub-determinant. If $\det A \neq 0$, then the smallest non-zero det is the whole NxN, so rank = N. If A has all zeros, then rank = 0.

Theorem R1: If the **rank** of an NxN matrix A is N-k, then there are only k eigenvectors of Ax=0.

Theorem R2: If a row in a matrix is linearly dependent on the other rows, then any segment of that row will be linearly dependent on the corresponding segments of the other rows.

Proof: If the whole row can be written as a lincom of other rows, this must be true of any piece of the row. If we say that $\mathbf{r_1} = \mathbf{r_2} + \mathbf{r_3}$ as vectors, then this sum must be true for each component and thus for any set of components, meaning for any subvector. And the rows are just such vectors.

Theorem R3: If the rank of an NxN matrix A is N-k, then k rows of A are linearly dependent on other rows. Only N-k rows of A are linearly independent. The same is true for the columns. This means that there are k "relations" among the rows, if k are linearly dependent on the others.

Suppose an NxN matrix has one row which is linearly dependent on the others. Then detA = 0 and we know that rank < N. Suppose two rows are linearly dependent. Any N-1 size submatrix you pick will have to include N-1 elements from at least one of these two rows. By the previous theorem, the (N-1)-length row segments in this submatrix will be linearly dependent, so detA' = 0 for all N-1 size submatrices. This means rank < N-1. And if 3 rows are linearly dependent, then detA" = 0 for all N-2 size submatrices, and so on. The same discussion can be repeated in terms of columns.

Here is another way to make the argument. Suppose all 3x3 sub dets are 0. Since this is true as we move horizontally on a specific set of 3 rows, this means that those 3 rows are not linearly indep. Thus, we cannot find ANY set of 3 rows that are lin indep. Thus, the number of lin indep rows must be < 3, and we say that rank < 3.

Theorem R4: $\dim(R_A)$ = rank [the dimensionality of the range in Ax=f is equal to the rank of A.]

Proof: The previous theorem says that N-Rank is the number of rows of a matrix that are linearly independent. The equation Ax = f is really a set of N-Rank equations in N unknowns. Suppose we have 2 independent equations in 3 unknowns x,y,z. Then we can just pick x and get (y,z) for that x. This means a solution like [x, y(x), z(x)] where x is anything. This is a "ray" in E^3. The range then has 1 dimension.
If we had only 1 equation in 3 unknowns, we could just pick x and y, then we would have [x, y, z(x,y)] and then our range would have 2 dimensions. These would be the axes in R_A:
[1,0,z(1,0)] and [0, 1, z(0,1)]. In the general case, we have N-Rank equations in N unknowns, so the dimension of R_A is going to be Rank.

Theorem R5: $R(A)^{\perp} = N(A^{\dagger})$. The perp-space of the range of A is equal to the nullspace of $A^{\dagger}$. (Alternative Theorem)

See page 153 Stakgold. Proof has two directions:

(a) If z is in $N(A^{\dagger})$ and f is in R(A), then $(f,z) = (Ax,z) = (x,A^{\dagger}z) = (x,0) = 0$. This says that z must therefore be in $R(A)^{\perp}$, so $N(A^{\dagger}) \subset R(A)^{\perp}$.

(b) If z is in $R(A)^{\perp}$ and f in R(A), then $0 = (f,z) = (Ax,z) = (x,A^{\dagger}z)$. Since x can be ANY point in the space, this means $A^{\dagger}z = 0$, so z must be in $N(A^{\dagger})$. Thus, $R(A)^{\perp} \subset N(A^{\dagger})$.

Theorem R6: For any NxN matrix A, rank(A) + nullity(A) = N.
Also, nullity(A) = nullity($A^\dagger$) and rank(A) = rank($A^\dagger$)

See Stakgold page 153-4. The conclusion of the previous theorem is that $N(A^\dagger) = R(A)^\perp$, so $\dim[N(A^\dagger)] = \dim[R(A)^\perp]$. We certainly know that $\dim[R(A)] + \dim[R(A)^\perp] = N$, so we know that $\dim[R(A)] + \dim[N(A^\dagger)] = N$. If we can just show that $\dim[N(A^\dagger)] = \dim[N(A)]$, which seems awfully reasonable, then $\dim[R(A)] + \dim[N(A)] = N$. This then says that rank(A) + nullity(A) = N.

This missing piece that $\dim[N(A^\dagger)] = \dim[N(A)]$. Stak shows this on page 154, along with the partner fact. Thus, he shows that nullity(A) = nullity($A^\dagger$) and rank(A) = rank($A^\dagger$). I will read this proof later.

Comments: This theorem really is very reasonable. Think about what happens as you apply A to all possible x in your space. For x in the nullspace subspace, all vectors map to 0 via Ax = 0. All vectors x in your space that are *not* in the nullspace must therefore map to some non-zero f in the range space via Ax = f. If the mapping Ax = f were somehow many-to-one, you could imagine that you might "lose" some dimensions through this mapping, and have dim(range) < N - dim(nullspace). But this does not happen and here is why. Suppose the nullspace has basis vectors e_1 and e_2 and there are two other basis vectors e_3 and e_4. Then let $Ae_3 = f_3$ and $Ae_4 = f_4$. If we had $f_3 = \alpha f_4$, then $A(e_3 - \alpha e_4) = 0$ and we would have $(e_3 - \alpha e_4)$ in the nullspace, which is impossible. Thus, the mapping from $N(A)^\perp \rightarrow R(A)$ does not "lose" any dimensions.