Hamilton-Cayley Theorem

1. Not every matrix is similar to a diagonal matrix, but every matrix is similar to a (possibly complex) upper triangular matrix. We prove this by induction. Obviously every 1 by 1 matrix is similar to an upper triangular matrix. Assume every n-1 by n-1 matrix is similar to an upper triangular matrix. Then let A be an n by n matrix; and let λ_1 be an eigenvalue, with eigenvector u_1. Then put u_1 in the first column of S, and pick the other columns of S to complete a basis for C^n, and then AS = SB, or $A = SBS^{-1}$ where B has the form:

$$\begin{bmatrix} \lambda_1 & w^T \\ 0 & B_1 \end{bmatrix}$$

   Since B_1 is an n-1 by n-1 matrix, $B_1 = P_1 T_1 P_1^{-1}$, where T_1 is an upper triangular n-1 by n-1 matrix. Then it can be verified that (remember that block matrices can be multiplied as if the blocks were just the elements, provided the block sizes are compatible):

$$\begin{bmatrix} \lambda_1 & w^T \\ 0 & B_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & P_1 \end{bmatrix} \begin{bmatrix} \lambda_1 & w^T P_1 \\ 0 & T_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & P_1^{-1} \end{bmatrix}$$

   or $B = PTP^{-1}$.

   Now since $A = SBS^{-1} = S(PTP^{-1})S^{-1} = (SP)T(SP)^{-1}$, A is similar to an upper triangular matrix.

2. If $p(T) = \sum_{i=0}^{n} \alpha_i T^i = 0$ for a certain matrix polynomial, and A is similar to T, then $p(A) = 0$. Proof: since $A = STS^{-1}$, $A^i = ST^iS^{-1}$, so $p(A) = \sum_{i=0}^{n} \alpha_i ST^iS^{-1} = Sp(T)S^{-1} = 0$.

3. (Hamilton-Cayley Theorem) If $p(\lambda)$ is the characteristic polynomial for A, then p(A) = 0. Proof: A is similar to an upper triangular matrix T, whose diagonal entries are the eigenvalues of T, and therefore also of A. Assume multiple eigenvalues are grouped together on the diagonal of T, so that T has the form:

$$\begin{bmatrix} T_1 & X & X & ... \\ 0 & T_2 & X & ... \\ 0 & 0 & T_3 & ... \\ ... & ... & ... & ... \end{bmatrix}$$

where T_i is an upper triangular block of size m_i by m_i, wheren m_i is the algebraic multiplicity of λ_i, and T_i has λ_i in each diagonal position.

Now the characteristic polynomial can be factored
$p(\lambda) = (\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2}(\lambda - \lambda_3)^{m_3}...$
So
$p(T) = (T - \lambda_1 I)^{m_1}(T - \lambda_2 I)^{m_2}(T - \lambda_3 I)^{m_3}...$

It can be verified by direct calculation that if T is a block triangular matrix then the diagonal blocks of T^m are just the m^{th} powers of the diagonal blocks of T. Thus $(T - \lambda_1 I)^{m_1}$ has the form:

$$\begin{bmatrix} (T_1 - \lambda_1 I)^{m_1} & X & X & ... \\ 0 & (T_2 - \lambda_1 I)^{m_1} & X & ... \\ 0 & 0 & (T_3 - \lambda_1 I)^{m_1} & ... \\ ... & ... & ... & ... \end{bmatrix}$$

Now $N = (T_1 - \lambda_1 I)$ is an m_1 by m_1 upper triangular matrix with 0s on the diagonal, and it can be verified that such a matrix is nilpotent, with $N^{m_1} = 0$. Thus $(T - \lambda_1 I)^{m_1}$ has the form:

$$\begin{bmatrix} 0 & X & X & ... \\ 0 & T_2' & X & ... \\ 0 & 0 & T_3' & ... \\ ... & ... & ... & ... \end{bmatrix}$$

Similarly, $(T - \lambda_2 I)^{m_2}$ has the form:

$$\begin{bmatrix} T_1' & X & X & ... \\ 0 & 0 & X & ... \\ 0 & 0 & T_3' & ... \\ ... & ... & ... & ... \end{bmatrix}$$

and $(T - \lambda_3 I)^{m_3}$ has the form:

$$\begin{bmatrix} T_1' & X & X & ... \\ 0 & T_2' & X & ... \\ 0 & 0 & 0 & ... \\ ... & ... & ... & ... \end{bmatrix}$$

4. The final step in the proof of Hamilton-Cayley is to show that $(T-\lambda_1 I)^{m_1}(T-\lambda_2 I)^{m_2}(T - \lambda_3 I)^{m_3}...$ is the zero matrix, so that p(T) = 0 and thus p(A) = 0. Notice that $(T - \lambda_1 I)^{m_1}$ is zero in the first column block (ie, first m_1 columns). Then you can verify by multiplying the block forms for $(T - \lambda_1 I)^{m_1}$ and $(T - \lambda_2 I)^{m_2}$ that this product is zero in the first two column blocks (ie, first $m_1 + m_2$ columns). Then verify that the product $(T-\lambda_1 I)^{m_1}(T-\lambda_2 I)^{m_2}(T-\lambda_3 I)^{m_3}$ is zero in the first three column blocks (first $m_1 + m_2 + m_3$ columns), and so on, so that the final product p(T) is zero in all columns.