The Jordan Form. For every integer $k > 0$, let J_k denote the $k \times k$-matrix obtained by augmenting the $(k-1)$-st identity matrix I_{k-1} by a trivial first column and last row (for $k = 1$, put $J_1 = 0$). Left multiplication by J_k shifts the entries of a column upward (losing the first one and replacing the last one by 0); similarly, right multiplication by J_k shifts the entries of a row to the right (losing the last one and replacing the first one by 0). Therefore, if $l < k$, we have the following equation of $k \times l$ - matrices:

$$J_k H_{kl} - H_{kl} J_l = E_{kl}, \tag{1}$$

where H_{kl} consists of the first l columns of J_k^T, and E_{kl} is zero except for a single 1 in the upper left corner.

All matrices to be considered will have entries in some fixed field F of scalars. The matrix $B_k(\lambda) = \lambda I_k + J_k$, with $\lambda \in F$, will be called the *Jordan block of degree k and eigenvalue λ*. We shall prove the following fact.

Theorem: *Every upper triangular $n \times n$-matrix A is similar to a direct sum of Jordan blocks.*

The proof will hinge on several applications of the obvious conjugation

$$\begin{bmatrix} I_r & -X \\ 0 & I_s \end{bmatrix} \begin{bmatrix} A' & 0 \\ 0 & A'' \end{bmatrix} \begin{bmatrix} I_r & X \\ 0 & I_s \end{bmatrix} = \begin{bmatrix} A' & A'X - XA'' \\ 0 & A'' \end{bmatrix} \tag{2}$$

where A' and A'' are square matrices of degrees r and s, respectively, and X is an $r \times s$ - matrix, with $n = r + s$. Let us write $U(X)$ for the $n \times n$-matrix with X in the upper right corner and zero elsewhere. Then (2) says: conjugating the direct sum $A = A' \oplus A''$ with $I_n + U(X)$ produces $A + U(A'X - XA'')$.

To prove the theorem, we imagine a counterexample A of minimal degree $n > 1$. Any conjugation $A \mapsto M^{-1}AM$ would still make a counterexample, and so would any translation $A \mapsto A + cI_n$. Therefore we may suppose that the entry a_{11} is zero, in other words, that the first column of A is trivial. Thus $A = (0 \oplus A^*) + U(R)$, where A^* is an $(n-1) \times (n-1)$ matrix, and R is an $(n-1)$-row. Now, the smaller matrix A^* *does* satisfy the theorem, and by conjugation on the last $n-1$ rows and columns, we can get $A^* = A_1 \oplus \cdots \oplus A_m$, where the A_i are Jordan blocks.

At this point we can write $A = (A' \oplus A'') + U(S)$, where $A'' = A_m$ and S is trivial below the first row. If we choose X in (2) with the same property, we get $A'X = 0$ from the trivial first column of A'. Thus, conjugating A with $I_n + U(X)$ yields $A = (A' \oplus A'') + U(S - XA'')$ in all (see?).

If $S - XA''$ could be made zero, A would not be a minimal counterexample. This eliminates any invertible A''. As we can arrange the Jordan blocks A_i in any order, we conclude that all their eigenvalues are zero, so that $A_i = J_{k(i)}$ for all i. Since the row-space of $A'' = J_s$ contains all vectors with trivial first coordinate, X can be chosen so as to make $S - XA'' = a_m E_{rs}$ for some $a_m \neq 0$.

By suitably permutating the last $n-1$ rows and columns, every one of the blocks A_i can be moved into the role of A'' and given the same treatment, until finally the first row of A equals $\big[0, R_1, \ldots, R_m\big]$, with $R_i = a_i E_{1,k(i)}$. For later, we arrange the indices so that $k(1) \geq k(i)$ for all i.

Multiplying the first row of A by $1/a_1$ (and its trivial first column by a_1), we now make $a_1 = 1$. This enlarges the Jordan block $A_1 = J_{k(1)}$ to $B' = J_r$, where $r = k(1) + 1$. Thus we can write our matrix as $B = (B' \oplus B'') + U(Q)$, where $B'' = A_2 \oplus \cdots \oplus A_m$. Partitioning the columns of the $r \times (n-r)$-matrix Q according to the same pattern, we get $Q = \big[Q_2, \ldots, Q_m\big]$, where $Q_i = a_i E_{r,k(i)}$. We shall kill Q by conjugating with a suitable $I_n + U(X)$.

If we let $X = \big[X_2, \ldots, X_m\big]$, such a conjugation will add $B'X - XB''$ to Q, that is, $J_r X_i - X_i J_{k(i)}$ in the columns corresponding to A_i. With $X_i = -a_i H_{r,k(i)}$, this expression yields exactly $-Q_i$, because of (1) and the fact that $r = k(1) + 1 > k(i)$ (see?). This demolishes Q and the counterexample.