Cayley Hamilton theorem: LinearAlg

Jonathan L.F. King
University of Florida, Gainesville FL 32611-2082, USA
squash@ufl.edu
Webpage http://squash.1gainesville.com/
17 November, 2017 (at 12:05)

Ques. Q1. Suppose two $\mathbf{F}$-matrices are conjugate over the algebraic closure of $\mathbf{F}$. Are they conjugate over $\mathbf{F}$? $\square$

Notation. Use $\wp(x) := \det(M - x \mathbf{I})$ for the characteristic poly of $M$. I'll use symbol $\simeq$ with the following meaning: Suppose $\wp$ is the characteristic polynomial of an $N \times N$ matrix $M$, or of a trn $\mathbf{T} : \mathbf{F}^N \rightarrow \mathbf{F}^N$, and $\mathbf{h}$ is a polynomial. I'll write $\wp \simeq \mathbf{h}$ to mean that

$$[-1]^N \cdot \wp = \mathbf{h}.$$ 

Use a similar convention for an alteration of the word “monic”: The phrase

"Consider a degree-$K$ monic polynomial $g \ldots"$

means that the high-order term of $g(x)$ is $\mbox{-}1^K x^K$.

Let $\mathbf{0}$ denote the zero-matrix or trn. Use $\tilde{\mathbf{0}}$ for the zero vector.

1: Cayley-Hamilton Theorem. Over field $\mathbf{F}$, consider an $N \times N$-matrix $M$. Then

$$\wp(M) = \mathbf{0}_{N \times N}.$$

So $M$ is a “root” of its own char-poly. $\diamond$

Proof when $M$ is upper-triangular. In matrix $M$, let $\alpha_1, \alpha_2, \ldots, \alpha_N \in \mathbf{F}$ be the diagonal entries; these are the eigenvalues of $M$. Using the std basis, let $\mathbf{E}_j := \mbox{Spn} \{(\mathbf{e}_1, \ldots, \mathbf{e}_j)\}$; so $\mathbf{E}_0 = \{\mathbf{0}\}$. Since $M$ is upper-triangular, the difference vector

$$\mathbf{d}_{j-1} := \mathbf{M} \mathbf{e}_j - \alpha_j \mathbf{e}_j$$

is in $\mathbf{E}_{j-1}$, for each $j \in [1..N]$. We want to show that each such $\mathbf{e}_j$ is annihilated by $\wp(M)$.

For $j \in [0..N]$, factor the characteristic polynomial as $\wp \simeq L_j \cdot R_j$, where the left & right are

$$L_j(x) := [x - \alpha_N] \cdot [x - \alpha_{N-1}] \cdot \ldots \cdot [x - \alpha_1];$$
$$R_j(x) := [x - \alpha_j] \cdot [x - \alpha_{j-1}] \cdot \ldots \cdot [x - \alpha_2] \cdot [x - \alpha_1].$$

[So $L_0 \simeq \wp$ and $R_0 = 1$.] All powers of $M$ mutually commute, thus

$$\wp(M) \simeq L_j(M) \cdot R_j(M).$$

Hence ISTShow that

$$Q[j]: \quad R_j(M) \text{ annihilates } \mathbf{E}_j.$$ 

Since all transformations annihilate $\mathbf{E}_0$, we need to prove $Q[j-1] \Rightarrow Q[j]$, for each $j = 1, 2, \ldots, N$.

Induction. Fix a $j \in [1..N]$ s.t $Q[j-1]$.

Firstly, $R_j(M)$ annihilates $\mathbf{e}_1, \ldots, \mathbf{e}_{j-1}$, since $R_{j-1}(M)$ does, and $R_j(M) = [M - \alpha_j \mathbf{I}] \cdot R_{j-1}(M)$. Secondly, to kill off $\mathbf{e}_j$ note that

$$R_j(M) \cdot \mathbf{e}_j = R_{j-1}(M) \cdot [M - \alpha_j \mathbf{I}] \cdot \mathbf{e}_j = R_{j-1}(M) \cdot \mathbf{d}_{j-1}.$$

This last product is $\mathbf{0}$, courtesy (2) and $Q[j-1].$ $\blacksquare$

Proof of C-H using JCF. We now handle a general $M$ by means of JCF, the Jordan Canonical Form thm. Let $\mathbf{G}$ denote the algebraic closure of $\mathbf{F}$. Viewing $M$ as acting on $\mathbb{G}^N$, our $M$ is conjugate (i.e similar) to its Jordan Canonical Form. Since the JCF is upper-triangular, the previous proof finishes the argument in the general case. $\diamond$

Elementary proof using a cyclic subspace. The preceding argument used two non-trivial theorems: JCFThm, as well as the result that a field has an algebraic closure.

Here is an elementary proof of C-H thm, never leaving field $\mathbf{F}$. Consider a trn $\mathbf{T}$ on a finite-dim’al $\mathbf{F}$-vectorspace and let $\wp$ be its characteristic poly. Fixing a vector $\mathbf{v}_0 \neq \mathbf{0}$, our goal is to show that

3: $\wp(\mathbf{T})(\mathbf{v}_0)$ equals $\mathbf{0}$.
Exer: Why does this suffice?
Iteratively define \( v_{j+1} := T(v_j) \) and stop at the first nullum \( N \) where \( T(v_N) \) is in the vectorspace
\[
W := \text{Spn}(v_0, v_1, \ldots, v_N).
\]
Define coeffs \( \alpha_j \) by
\[
4: \quad T(v_N) := \sum_{j+k=N} \alpha_j v_k,
\]
where such sums are taken over nullums \( j \) and \( k \).
Notice that our \( W \) is an \( \mathcal{T} \)-invariant subspace. And the linearly-independent (exercise!) tuple
\[
\mathcal{B} := (v_0, \ldots, v_N)
\]
is a basis for subspace \( W \).

5: Companion Lemma. With \( \mathcal{T}, W \) and \( \mathcal{B} \) as above, let \( M \) be the \([N+1] \times [N+1]\) matrix of \( T \) relative to ordered \( W \)-basis \( \mathcal{B} \). Then
\[
5a: \quad M = \begin{bmatrix}
0 & 0 & \cdots & \alpha_N \\
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1
\end{bmatrix}
\]
And its characteristic polynomial is
\[
5b: \quad \varphi_M(x) \cong x^{N+1} - \sum_{j+k=N \text{ with } j,k \in \mathbb{N}} \alpha_j x^j .
\]

Remark. A matrix of form (5a) is a companion matrix. It is “the companion matrix of polynomial (5b)”. Wikipedia has a nice write-up. \( \Box \)

Proof of (5). The \([N+1] \times [N+1]\) matrix \( xI - M \) is
\[
\begin{bmatrix}
x & -\alpha_N \\
-1 & x & -\alpha_{N-1} \\
-1 & x & -\alpha_{N-2} \\
\vdots & \vdots & \ddots \\
-1 & x & -\alpha_3 \\
-1 & x & -\alpha_2 \\
-1 & x & -\alpha_1 \\
-1 & [x - \alpha_0]
\end{bmatrix}
\]
We compute its determinant by summing products over transversals. The main diagonal yields
\[
\hat{\tau}_0: \quad x^N \cdot [x - \alpha_0] \overset{\text{note}}{=} x^{N+1} - \alpha_0 x^N.
\]
Now, in columns 0, 1, \ldots, \( N-1 \) we either choose “\( x \)” or “-1”. In a column where we choose -1, the row of our choice prevents us from choosing \( x \) in the next column; so we must again choose -1. Thus:

Once we leave the main diagonal, we must stay on the first off-diagonal.

So what is the contribution to \( \text{Det}(xI - M) \) from a transversal with \( j \in [1..N] \) many -1’s? It is
\[
x^{N-j} \cdot [-1]^j \cdot [-\alpha_j] \cdot \text{[Sign of permutation]} .
\]
The sign of the perm is \([-1]^j \), so the \( j \)-th-transversal contribution to \( \varphi_M(x) \) is
\[
\hat{\tau}_j: \quad -[\alpha_j \cdot x^{N-j}] .
\]
Adding (\( \hat{\tau}_0 \)) to \( \sum_{j=1}^N (\hat{\tau}_j) \) yields \( \text{Rhs}(5b) \). \( \blacksquare \)

Second Proof of C-H. The given trm \( T \) and vector \( v_0 \) determine a \( \mathcal{T} \)-invariant subspace \( W \) and matrix \( M \), as above. An easy exercise – see the Triangular Matrix Lemma in the \text{jordan_decomp} latex file – shows that \( \varphi_M \) is a factor-poly of \( \varphi_T \). So (3) will follow from showing that \( \varphi_M(T) \) annihilates \( v_0 \). And this follows from (5b) and (4). \( \blacksquare \)

6: Corollary. Fix \( K \in \mathbb{Z}_k \) and an arbitrary degree-\( K \) \( m \)-nic \( F \)-poly \( g() \). Then there exists a \( K \times K \) matrix over \( F \) whose characteristic-poly equals \( g \).
\[\text{Pf.} \] Use matrix (5a) with \( K := N+1 \).
\( \blacksquare \)

7: Application. Let \( F := \mathbb{Z}_p \), where \( p \) is prime. To produce a \( p \times p \) \( F \)-matrix \( M \) with no \( F \)-eigenvalues, pick a non-zero element \( \beta \in F \), and define
\[
g(x) := \beta + \prod_{\gamma \in F} [x - \gamma].
\]
For instance, consider \( p = 3 \) and \( \beta = -1 \). Then
\[
g(x) = -1 + x[x-1][x+1] = x^3 - [x+1] = x^3 - [\alpha_0 x^2 + \alpha_1 x + \alpha_2],
\]
using the notation of (5b), where \( \alpha_0 := 0, \alpha_1 := 1 \) and \( \alpha_2 := 1 \). Courtesy our (5a), then, matrix
\[
M := \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}
\]
has no eigenvalues in \( \mathbb{Z}_3 \).

End Notes

First, we need a general lemma.

8: Lemma. Fix fields \( \mathbb{G} \supset \mathbb{F} \) and consider a collection \( \mathcal{C} \subset \mathbb{F}^{\times N} \) of vectors which is linearly dependent over \( \mathbb{G} \). (Typically, \( \mathbb{G} \) is the algebraic closure of \( \mathbb{F} \).) Then \( \mathcal{C} \) is already linearly dependent over \( \mathbb{F} \).

Proof. View \( \mathbb{F}^{\times N} \)-vectors as column vectors, and use \( \vec{0} \) for the col-vec of all zeros. FTSOC, suppose we have a non-trivial dependence
\[
\sum_{j=1}^{7} \alpha_j \cdot \vec{c}_j = \vec{0},
\]
for scalars \( \alpha_j \in \mathbb{G} \) and col-vecs in \( \vec{c}_j \in \mathcal{C} \). Some \( \alpha_j \neq 0 \), so WLOG \( \alpha_1 \neq 0 \). By multiplying (\( \dagger \)) by \( 1/\alpha_1 \), WLOG \( \alpha_1 = 1 \).

Now view \( \mathbb{G} \) as an \( \mathbb{F} \)-vectorspace. Collection \( \{1\} \) is L.I, so the axiom-of-choice says we can extend it to to a \( \mathbb{F} \)-basis \( \{1\} \uplus \mathcal{E} \) for \( \mathbb{G} \). So \( \mathcal{E} \subset \mathbb{G} \), and every \( \alpha \in \mathbb{G} \) can be uniquely written as an \( \mathbb{F} \)-linear combination of “vectors” in \( \mathbb{G} \).

Define a linear map \( \text{Proj} : \mathbb{G} \rightarrow \mathbb{F} \) by \( 1 \mapsto 1 \) and, for each \( \vec{e} \in \mathcal{E} \), have \( \text{Proj} \) send \( \vec{e} \mapsto 0 \). Whence \( \text{Proj}(\cdot) \) is the identity on \( \mathbb{F} \), and for \( \alpha, \beta \in \mathbb{G} \) and \( f \in \mathbb{F} \):
\[
\text{Proj}(\alpha + \beta) = \text{Proj}(\alpha) + \text{Proj}(\beta); \quad \text{Proj}(\alpha \cdot f) = \text{Proj}(\alpha) \cdot f.
\]

Applying map \( \text{Proj}^{\times N} : \mathbb{G}^{\times N} \rightarrow \mathbb{F}^{\times N} \) to (\( \dagger \)) yields
\[
\sum_{j=1}^{7} \text{Proj}(\alpha_j) \cdot \vec{c}_j = \vec{0}
\]
by (\( \ast \)), since each entry in each \( \vec{c}_j \) is in \( \mathbb{F} \).

And \( \text{Proj}(\alpha_1) = \text{Proj}(1) = 1 \), which is not zero. So (\( \dagger \)) exposes a non-trivial \( \mathbb{F} \)-linear-dependence of \( \mathcal{C} \).

Minimal poly of \( M \)

See jordan_decomp.latex for theorems used below. The minimal polynomial of an \( \mathbb{F} \)-matrix \( M \) is the smallest-degree monic \( \mathbb{F} \)-polynomial \( \mathcal{Y} \) such that \( \mathcal{Y}(M) = 0 \). Applying (8) to collection \( \mathcal{C} := \{M^j\}_{j \in \mathbb{N}} \) shows, if we take the smallest-degree monic \( \mathbb{G} \)-polynomial, that we still get \( \mathcal{Y} \).

For a \( \lambda \in \mathbb{G} \), consider the \( D \times D \) Jordan Block
\[
J := \lambda \cdot \text{J}B(D) := \begin{bmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \ddots & 1 \\ & & & \lambda & 1 \end{bmatrix}.
\]

It is the sum \( \lambda I + N \), where \( N \) is the nilpotent matrix \( 0 \cdot \text{J}B(D) \). For \( R \in \mathbb{N} \), the Binomial thm applies, since \( N \approx I \), to the \( R \)-th power of \( J \) to assert
\[
J^R = \sum_{j+k=R} \lambda^j \cdot \binom{R}{j,k} \cdot N^k.
\]

For \( R \in [0..D) \), then, \( J^R \) has 1’s on the \( R \)-th off-diagonal, and 0’s on all higher diagonals. Thus \( \{I, J, \ldots, J^{D-1}\} \) is a lin-indep collection of matrices. And \( J^D = \lambda^D I \). So \( \text{Deg}(\mathcal{Y}_J) \) equals \( D \). Therefore,
\[
\mathcal{Y}_J(x) \bowtie \mathcal{Y}_J(x) = [x - \lambda]^D.
\]

9: Fact. The characteristic and minimum polynomials satisfy
\[
\mathcal{Y}_M = \mathcal{Y}_A \cdot \mathcal{Y}_B ; \quad \mathcal{Y}_M = \text{Lcm}(\mathcal{Y}_A, \mathcal{Y}_B),
\]
when \( M := \text{Diag}(A,B) \) is a block-diagonal matrix. Proof. Immediate.
**Caveat.** Suppose $M$ is block upper-triangular; it has square-blocks $B_1, \ldots, B_L$ along the diagonal, zeros south-west of these blocks, and possibly non-zero values north-east of these blocks. Certainly

$$\varphi_M = \varphi_{B_1} \cdot \varphi_{B_2} \cdot \ldots \cdot \varphi_{B_L}.$$ 

However, the corresponding stmt for $\Upsilon_M$ with Lcm is false.

As a CEX, the matrices $[0 \ 0]$ and $[0 \ 1]$ have the same $1 \times 1$ diagonal-blocks, and the same char-poly, but different min-polys; they are $x$ and $x^2 \, . \, □$

An eigenvalue is a "simple" eigenvalue if its eigenspace is 1-dim’al.

**10: Corol.** A block-diagonal $M$ has "equality" $\varphi_M \simeq \Upsilon_M$ IFF $M$ has only simple eigenvalues. ◊

**Proof.** This follows from either (9) or (12). ◆

A downturn $\vec{D}$ is a sequence of positive integers $D_1 \geq D_2 \geq \ldots \geq D_\varepsilon$. It yields the JCF

$$\lambda_{\cdot \cdot \cdot \cdot} := \text{Diag} \left( \lambda_{\cdot \cdot \cdot \cdot}(D_1), \ldots, \lambda_{\cdot \cdot \cdot \cdot}(D_\varepsilon) \right)$$

of the general $\lambda$-nilpotent matrix. Use Size($\vec{D}$) for $D_1 + \ldots + D_\varepsilon$.

For $F$-matrix $M$, suppose that $\lambda_1, \ldots, \lambda_L$ are the distinct $G$-eigenvalues. The eigenvalues yield a unique list $\vec{D}^1, \vec{D}^2, \ldots, \vec{D}^L$ of downtups (of varying lengths) so that

**11: Theorem.** With $F$-matrix $M$ having JCF (11), its characteristic and minimum polys are

$$\varphi_M(x) \simeq \prod_{\ell=1}^L [x - \lambda_\ell]^\text{Size}(\vec{D}_\ell) \quad \text{and} \quad \Upsilon_M(x) = \prod_{\ell=1}^L [x - \lambda_\ell]^\text{Max}(\vec{D}_\ell).$$

Necessarily, these polynomials have all their coefficients in $F$. **Proof.** Use (8) and (9). ◊

**13: Fact.** Suppose $M$ is an $N \times N$ matrix. Write its char and min polys as

$$\varphi_M(x) \simeq [x - \lambda_1] \cdot \ldots \cdot [x - \lambda_N] \quad \text{and} \quad \Upsilon_M(x) = [x - \beta_1] \cdot \ldots \cdot [x - \beta_D].$$

[The $\beta$’s form a sub-multiset of the $\lambda$’s.] Then for each nz-scalar $\sigma$:

$$\varphi_{\sigma M}(x) = \sigma \varphi_{M}(\frac{1}{\sigma}x),$$

and analogously for the min-poly. ◊

**Continuity.** Easily, the mapping $M \mapsto \varphi_M$ iscts. But neither $\Upsilon_M$ nor JCF($M$) varies continuously with $M$. For $\beta \neq 0$, define $7 \times 7$ matrices

$$M_\beta := \begin{bmatrix} 0 & \beta & & & & & \\ \vdots & \ddots & \ddots & & & & \\ & \vdots & \ddots & \ddots & \ddots & & \\ & & \vdots & \ddots & \ddots & \ddots & \\ & & & \vdots & \ddots & \ddots & \ddots \\ & & & & \vdots & \ddots & \ddots \\ & & & & & \ddots & \ddots \\ & & & & & & 0 \end{bmatrix} \quad \text{and} \quad J := \begin{bmatrix} 0 & 1 & & & & & \\ & 0 & \ddots & & & & \\ & & \ddots & \ddots & & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \ddots & \ddots & \ddots \\ & & & & & \ddots & \ddots \\ & & & & & & 0 \end{bmatrix}.$$ 

The JCF (Jordan Canonical Form) of $M$ is $J$. So the min-poly $\Upsilon_M(x) = \Upsilon_J(x) = x^7$. But as $\beta \to 0$, our $M_\beta$ goes to $0_{7 \times 7}$, whose min-poly is $x$. This example also shows that neither eigenspaces nor nilspaces vary ctsly.

**Invariant properties.** Suppose $S$ is invertible. Since $ST$ is conjugate (exercise) to $TS$, they have the same min-poly and char-poly. We now generalize char-poly to non-invertible:

**14: Lemma.** For $S, T \in \text{MAT}_{N \times N}(F)$: Products $ST$ and $TS$ have the same characteristic poly. ◊

---

\(^{\cdot}1\) INDIRECT: The only $M$ eval is 0, yet Rank($M$) = 7 − 1. So the nullspace, i.e 0-espace, is only 1-dimensional, hence JCF($M$) has only one JB.

DIRECT: Consider ordered-basis $V := \{ v_1, \ldots, v_7 \}$, where we define $v_k := \beta^j e_k$ with $j + k = 7$. The lefthand action of $M$, when expressed w.r.t $V$, is $J$. Equivalently, $J = C^{-1}MC$ where $C$ is the diagonal matrix with entries $\beta^6, \beta^5, \ldots, \beta, 1$. 

Filename: Problems/Algebra/LinearAlg/cayley_hamilton.lateX
Proof. We can proceed as follows if $F$ has a topology, with the field operations cts, so that $GL(F^{\times N})$ is dense in $Lin(F^{\times N})$. For then, take invertible matrices $S_j$ which converge to $S$ and use that the char-poly varies continuously. ♦

Here is a standard “Algebraist’s argument” Let $\tilde{F}$ be the field generated by $F$ and $N^2$ independent transcendentals. Let $\tilde{S}$ be a matrix obtained by putting a distinct transcendental in each position.

Since $\tilde{S}$ is $\tilde{F}$-invertible, $\tilde{S}T$ and $TS$ have the same char-poly. Now apply the ring-hom $\varphi:F\rightarrow F$ which sends each transcendental to its corresponding $F$-element in $S$. (I.e, plug in the $S$-values for the corresponding transcendentals in $\tilde{S}$.) ♦

Note. We used that the above ring-hom $\varphi:\tilde{F}\rightarrow F$ preserves determinants (since it preserves mult and addition) hence preserves charpolys.

However, this argument does not show that $ST$ is conjugate to $TS$. Why? Well, $\varphi$ can carry an invertible matrix to a non-invertible. Perhaps $\varphi$ carries every matrix conjugating $\tilde{S}T$ to $TS$, to a non-invertible puppy.

Here is an example: Let $S := \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$, $T := \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. Then $ST$ is the zero-matrix, but $TS$ equals $S$. So not only is $ST$ not similar (not conjugate to) $TS$, they even have different minpolys, hence different JCF's. Since $S$ is the limit of $S_x := \begin{bmatrix} x & 0 \\ 1 & 0 \end{bmatrix}$ as $x \downarrow 0$, we have another example showing that the minimum-poly and JCF do not vary ctsly. □