$A$ を $n$ 次正方行列、$A$ の固有多項式を $p(\lambda) = |A - \lambda I|$、また $A - \lambda I$ の adjugate matrix を $B$ とすると

$$p(\lambda)I = B(A - \lambda I)$$

が成り立つ。$p(\lambda)$ は $\lambda$ の $n$ 次式、$B$ の各要素は $\lambda$ の $n-1$ 次以下の多項式となるので、例えば $n = 3$ のとき、$p(\lambda) = p_0 + p_1\lambda + p_2\lambda^2 + p_3\lambda^3$、$B = B_0 + B_1\lambda + B_2\lambda^2$ と置くことができ、上の等式は

$$p_0 I + (p_1 I)\lambda + (p_2 I)\lambda^2 + (p_3 I)\lambda^3 = (B_0 + B_1\lambda + B_2\lambda^2)(A - I\lambda) \cdots\cdots (*)$$

と書ける。いま、この等式の左辺・右辺の $\lambda$ を行列 $X$ に置き換えた式を考え、

$$p(X) = p_0 I + (p_1 I)X + (p_2 I)X^2 + (p_3 I)X^3$$

と

$$f(X) = (B_0 + B_1 X + B_2 X^2)(A - IX)$$

とを比較する。最終的に証明したいこと (Cayley-Hamilton の定理) は $p(A) = O$ である。もしも、スカラー $\lambda$ に対する (*) 同様、任意の行列 $X$ に対して $p(X) = f(X)$ が成り立ってくれるなら、直ちに $p(A) = f(A) = O$ が示される。しかし残念ながらそう簡単には行かない。

$f(X)$ を展開すると

$$f(X) = B_0 A + (-B_0 X + B_1 XA) + (-B_1 X^2 + B_2 X^2 A) + (-B_2 X^3)$$

となるが、必ずしも $XA = AX$ ではないために

$$g(X) = B_0 A + (-B_0 + B_1 A)X + (-B_1 + B_2 A)X^2 + (-B_2)X^3$$

と等しくなるとは限らない。しかし $g(A)$ を計算するとやはり $O$ となるので、$p(A) = g(A)$ が示されれば定理は証明されたことになる。

ここで、$f(A)$ だけでなく $g(A)$ も $O$ になることは、もちろん直接代入して計算すれば分かることであるが、$X = A$ という状況が、まさに $XA = AX$ を満たすために、$f(X) = g(X)$ を成立させることからも頷ける。

さて、実は $p(X)$ は $f(X)$ とは必ずしも等しくならない一方で、$g(X)$ とは任意の $X$ において等しくなる。そのおかげで $p(A) = g(A) = O$ が示されるのである。以下、このことを証明する。

いま $X = \lambda I$ の場合を考えると、

$$\begin{aligned} p(\lambda I) &= p_0 I + (p_1 I)(\lambda I) + (p_2 I)(\lambda I)^2 + (p_3 I)(\lambda I)^3 \\ &= p_0 I + (p_1 I)\lambda + (p_2 I)\lambda^2 + (p_3 I)\lambda^3 \end{aligned}$$

これは (*) の左辺と同じである。一方

$$\begin{aligned} g(\lambda I) &= B_0 A + (-B_0 + B_1 A)(\lambda I) + (-B_1 + B_2 A)(\lambda I)^2 + (-B_2)(\lambda I)^3 \\ &= B_0 A + (-B_0 + B_1 A)\lambda + (-B_1 + B_2 A)\lambda^2 + (-B_2)\lambda^3 \end{aligned}$$

は (*) の右辺を展開したものに等しく、$p(\lambda I) = g(\lambda I)$ となる。すなわち、$X = \lambda I$ のときに限って $p(X) = g(X)$ が示されたが、これは $\lambda$ に関わらず成り立つことから、$X$ の各次数について、$p(X)$ と $g(X)$ の係数（行列）は一致していなければならない。すると結局、任意の $X$ について $p(X) = g(X)$ であることが分かる。