Apprentice Program

Miklos Abert
Notes taken by Matt Holden

July 10, 2006

Lemma 1. *A polynomial of degree n is uniquely determined by its value at $n+1$ distinct points $x_0, \ldots, x_n$.*

Proof. Suppose f and g are degree n polynomials such that $f(x_i) = g(x_i)$ for $i = 0, \ldots, n$. Then $f - g$ has degree $\leq n$ but it has $n+1$ roots (the x_i), so it must be the zero polynomial, hence $f \equiv g$. $\square$

We now consider the question of the *existence* of a polynomial attaining prescribed values at the $n+1$ points. Suppose we are given scalars $y_0, \ldots, y_n$ and we want to find a polynomial f such that $f(x_i) = y_i$ for all i. For each fixed $i = 0, \ldots, n$, notice that the polynomial

$$f_i(x) = \prod_{j \neq i} \frac{x - x_j}{x_i - x_j}$$

satisfies $f_i(x_i) = 1$ and $f_i(x_j) = 0$ for all $j \neq i$. Thus, we consider the polynomial

$$f(x) = \sum_{i=0}^{n} y_i \prod_{j \neq i} \frac{x - x_j}{x_i - x_j}, \tag{1}$$

which clearly satisfies $f(x_i) = y_i$ for all i. Since we already proved uniqueness, we have the following:

Theorem 2 (Lagrange interpolation). *Given any $n+1$ distinct points $x_0, \ldots, x_n$ and any scalars $y_0, \ldots, y_n$, there exists a unique polynomial f (given by the formula (1)) such that $f(x_i) = y_i$ for all i.*

The characteristic polynomial

If $M \in M_n(k)$ then the $n^2 + 1$ matrices $I, M, \ldots, M^{n^2}$ cannot be linearly independent over k, since $M_n(k)$ is a k-vector space of dimension n^2. Thus, we can find scalars $a_0, \ldots, a_{n^2} \in k$ such that

$$a_0 I + a_1 M + \cdots + a_{n^2} M^{n^2} = 0.$$

This shows that every $n \times n$ matrix M is a root of a polynomial of degree at most n^2. Our next goal is to prove the following:

Claim 3. *Every $M \in M_n(k)$ is a root of a degree n polynomial.*

To see that this bound is sharp, consider the matrix with 1s on the first super-diagonal:

$$M = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & & \ddots & 1 \\ 0 & \cdots & & 0 \end{pmatrix}$$

For $k = 1, \ldots, n-1$, M^k is the matrix with 1s on the kth super-diagonal (you should check this), and $M^n = 0$. It follows that $I, M, \ldots, M^{n-1}$ are linearly independent (over k), so M cannot satisfy a polynomial of degree $< n$. Thus, M is a root of the polynomial x^n, but no polynomial of lower degree, so the bound in the claim is sharp.

Definition 4. The *characteristic polynomial* of a matrix $M \in M_n(k)$ is

$$\mathrm{ch}_M(\lambda) = \det(\lambda I - M),$$

which is a degree n polynomial in the variable λ, with coefficients in k.

Notice that $\mathrm{ch}_M(\lambda)$ has leading coefficient 1, so we can write it as

$$\mathrm{ch}_M(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0. \tag{2}$$

Claim 3 will follow from the following important result:

Theorem 5 (Cayley-Hamilton). *M is a root of* $\mathrm{ch}_M(\lambda)$.

Before beginning the proof, recall how we defined the quasi-inverse of a square matrix. Given a matrix A, we define a new matrix B by setting $b_{ij} = (-1)^{i+j} \det A_{ji}$, and we showed that $AB = (\det A)I$. In particular, $B = (\det A)A^{-1}$ if A is invertible.

Proof of theorem. Define a matrix B by setting

$$b_{ij} = (-1)^{i+j} \det \big[(\lambda I - M)_{ji}\big],$$

and notice that

$$B(\lambda I - M) = \det(\lambda I - M) \cdot I = \mathrm{ch}_M(\lambda) \cdot I. \tag{3}$$

Now, B is a matrix of polynomials of degree $\leq n-1$, so we can write

$$B = B_{n-1}\lambda^{n-1} + \cdots + B_1\lambda + B_0, \tag{4}$$

where $B_{n-1}, \ldots, B_0$ are constant scalar matrices. (This is just the natural isomorphism $M_n\big(k[\lambda]\big) \cong M_n(k)[\lambda]$: "a polynomial matrix equals a matrix polynomial".) Substituting (2) and (4) into (3) yields

$$(B_{n-1}\lambda^{n-1} + \cdots + B_1\lambda + B_0)(\lambda I - M) = (\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0)I.$$

Equating coefficients of λ gives a system of $n+1$ equations:

$$\begin{aligned} -B_0 M &= a_0 I \\ -B_1 M + B_0 &= a_1 I \\ &\vdots \\ -B_k M + B_{k-1} &= a_k I \\ &\vdots \\ -B_{n-1} M + B_{n-2} &= a_{n-1} I \\ B_{n-1} &= I. \end{aligned}$$

Multiply these equations by $I, M, \ldots, M^n$, respectively, and add them. The RHS of this sum equals $a_0 I + a_1 M + \ldots a_{n-1}M^{n-1} + M^n = \mathrm{ch}_M(M)$, while the LHS telescopes to zero:

$$(-B_0 M) + (-B_1 M + B_0) \cdot M + \cdots + B_{n-1} \cdot M^n = 0,$$

hence $\mathrm{ch}_M(M) = 0$, as claimed. $\square$

Definition 6. Let $\phi \in \mathrm{Hom}(V, V)$ be a linear transformation of a vector space V. We say that $0 \neq v \in V$ is an *eigenvector* of ϕ if $\phi(v) = \lambda v$ for some $\lambda \in k$. We then say that λ is the *eigenvalue* associated to v.

Given any $\lambda \in k$, we define $V_\lambda = \{v \in V : \phi(v) = \lambda v\}$. It is easy to check that V_λ is a subspace of V. It is a nonzero subspace iff λ is an eigenvalue of ϕ, in which case we call V_λ the *eigenspace* associated to λ. Finally, notice that if $\lambda \neq \mu$ then $V_\lambda \cap V_\mu = \{0\}$.