f(A) Part 1: Diagonalizable A

Suppose $f : \mathbb{R} \to \mathbb{R}$ is continuous. We investigate here how it makes sense to "extend" f to a function $f : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ which is a continuous function of the entries of a square matrix. While f in the former sense is a different kind of beast than in the latter sense, but we will use f do denote them both, nonetheless.

We consider here the generic (and easier) case: assume A is diagonalizable. Then there is an invertible $V \in \mathbb{C}^{n \times n}$ such that $AV = V\Lambda$ and

$$V^{-1}AV = \Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & 0 & \lambda_{n-1} & 0 \\ 0 & 0 & \dots & 0 & \lambda_n \end{bmatrix} \equiv \text{diag}(\lambda_i).$$

Let p be a polynomial $p(x) = a_0 + a_1 x + \cdots + a_d x^d$ and define $p(A)$ as $p(A) = a_0 I + a_1 A + \cdots + a_d A^d$. Then

$$\begin{aligned} p(A) = p(V\Lambda V^{-1}) &= a_0 I + a_1 V\Lambda V^{-1} + \cdots + a_d (V\Lambda V^{-1})^d \\ &= V(a_0 I + a_1 \Lambda + \cdots + a_d \Lambda^d)V^{-1} \\ &= Vp(\Lambda)V^{-1} \\ &= V \ \text{diag}(p(\lambda_i)) \ V^{-1}. \end{aligned}$$

Now any continuous function can be arbitrarily closely approximated over any finite interval by a polynomial (Weierstrass Approximation Theorem), so this definition for $p(A)$ will essentially fix our definition for $f(A)$ [this is a *continuity argument*, which basically says that since polynomials are arbitrarily close to any continuous function f, the definition of $f(A)$ must generalize $p(A)$, otherwise $f(A)$ would not be a continuous function of the the entries of A]. Therefore however we define $f(A)$, it must satisfy $f(A) = Vf(\Lambda)V^{-1}$. But for diagonalizable matrices this is the whole story [the eigenvalues A are continuous functions of the entries of A], for the only definition which generalizes $p(A)$ is

$$f(A) = f(V\Lambda V^{-1}) = Vf(\Lambda)V^{-1} = Vf(\text{diag}(\lambda_i))V^{-1} = V \ \text{diag}(f(\lambda_i)) \ V^{-1}.$$

Notice that $f(A)$ depends only on $f(\lambda_i), \ i = 1{:}n$, and in fact $f(A) = g(A)$ for any functions f and g such that $f(\lambda_i) = g(\lambda_i), \ i = 1{:}n$. But wait! That means $f(A) = p(A)$ for any polynomial interpolator of f on the nodes $\lambda_i, \ i = 1{:}n$. Specifically, $f(A) = p_{A,f}(A)$, where $p_{A,f}$ is the interpolating polynomial of degree $n-1$ or less for $(\lambda_i, f(\lambda_i)), \ \ i = 1{:}n$.

This is almost the whole story for diagonalizable A; but there is a snag. Do you see the problem with our definition? $A \in \mathbb{R}^{n \times n}$ may have non-real (complex) eigenvalues, but we said $f : \mathbb{R} \to \mathbb{R}$, so it may not make sense to talk about $f(\lambda_i)$ if $\lambda_i \notin \mathbb{R}$. We will take the easy way out of this dilemma, by requiring that $f : \mathbb{C} \to \mathbb{C}$ makes sense. We will need to restrict f again as we consider nondiagonalizable matrices...