Consequences and applications of the Jordan theorem

Functions of an operator

Suppose that $\underline{A}\colon \mathcal{V} \to \mathcal{V}$ is an operator and $f\colon \mathbf{C} \to \mathbf{C}$ is an ordinary numerical function. (In fact, the domain of f might be only a subset of $\mathbf{C}$.) Does

$$f(\underline{A})$$

have any meaning? (Of course, we would not raise this question if it did not have a *useful*, positive answer.)

Let's look at some cases where the answer is already *yes*:

$\underline{A}^2 \equiv \underline{A} \circ \underline{A} \;\Rightarrow\;$ polynomial functions of $\underline{A}$ are defined.

$\dfrac{1}{\underline{A}} \equiv \underline{A}^{-1} \;\Rightarrow\;$ rational functions are defined. (The value of the function may not exist for all $\underline{A}$.) Example:

$$\frac{\underline{A}-i}{\underline{A}+i} \equiv (\underline{A}+i)^{-1}(\underline{A}-i) = (\underline{A}-i)(\underline{A}+i)^{-1}$$

(if $(\underline{A}+i)^{-1}$ exists).

Remark: $\dfrac{\underline{A}-i}{\underline{B}+i}$ is not defined if $[\underline{A},\underline{B}] \neq \underline{0}$.

Observe:

1) For a diagonal matrix $A = D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$,

$$f(D) = \begin{pmatrix} f(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & f(\lambda_n) \end{pmatrix}$$

for rational f.

2) Since $f(\underline{A})$ depends only on the *operator* $\underline{A}$, not on the matrix (choice of basis), we must have

$$f(SAS^{-1}) = S\,f(A)\,S^{-1}$$

under any similarity transformation. To verify this for polynomials, note that internal S's cancel out: e.g.,

$$(SAS^{-1})^2 = SAS^{-1}\,SAS^{-1} = SAAS^{-1} = S\,A^2\,S^{-1}.$$

This suggests the following **Idea**: For a general f (say $f(x) \equiv e^x$), define $f(D)$ for a diagonal matrix D as

$$\begin{pmatrix} f(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & f(\lambda_n) \end{pmatrix}.$$

Then for a diagonalizable A, $\ A = SDS^{-1}$, define

$$f(A) \equiv S\,f(D)\,S^{-1}.$$

Thus $f(\underline{A})$ is well-defined.

In other words, for $\underline{A} = \sum_{\nu=1}^{L} \lambda_\nu \underline{P}_\nu$, define $f(\underline{A}) \equiv \sum_{\nu=1}^{L} f(\lambda_\nu)\underline{P}_\nu$.

Let's check this for polynomials:

$$\underline{A}^2 = \left(\sum_\nu \lambda_\nu \underline{P}_\nu\right)\left(\sum_\mu \lambda_\mu \underline{P}_\mu\right) = \sum_\nu {\lambda_\nu}^2 \underline{P}_\nu$$

since ${\underline{P}_\nu}^2 = \underline{P}_\nu$ and $\underline{P}_\nu \underline{P}_\mu = 0$ if $\nu \neq \mu$. [What happens if f is the characteristic function of a set?]

Some other definitions of $f(\underline{A})$ are possible (for one who knows enough analysis):

(1) *Power series.* Examples:

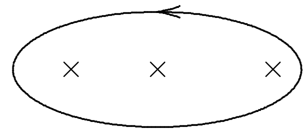
$$e^{\underline{A}} \equiv \underline{1} + \underline{A} + \frac{1}{2}\,\underline{A}^2 + \cdots = \sum_{n=0}^{\infty} \frac{1}{n!}\,\underline{A}^n\,;$$

$$\ln\,(\underline{1} - \underline{A}) \equiv -\sum_{n=1}^{\infty} \frac{1}{n}\,\underline{A}^n \qquad \text{— or} \qquad \ln \underline{A} \equiv -\sum_{n=1}^{\infty} \frac{1}{n}(\underline{1} - \underline{A})^n.$$

If the series converges (see homework), this agrees with the diagonalization definition, where *it* is applicable. [Check this.]

(2) *Cauchy's formula.* This requires f to be analytic, but there is no restriction on the radius of convergence.

$$f(\underline{A}) \equiv \frac{1}{2\pi i} \oint \frac{f(z)}{z-\underline{A}}\, dz$$



for any contour surrounding all eigenvalues of $\underline{A}$ in the usual way. This equation means that for all $\vec{v}$,

$$f(\underline{A})\vec{v} = \frac{1}{2\pi i} \oint [f(z)\,(z-\underline{A})^{-1}\vec{v}]\, dz.$$

As a homework problem, I ask you to check that this definition is consistent with diagonalization.

If $\underline{A}$ is *not diagonalizable*, let's assume that f is analytic and define $f(\underline{A})$ by power series. (The Cauchy definition could also be used.) For $\vec{v} \in \mathcal{U}(\lambda_\nu)$, we'll expand $f(x)$ about $x = \lambda_\nu$:

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{n!}\, f^{(n)}(\lambda_\nu)\,(x-\lambda_\nu)^n.$$

Then, for x, substitute $\underline{A}\big|_{\mathcal{U}(\lambda_\nu)} \equiv \underline{A}_\nu$ to find $f(\underline{A})\vec{v}$. (Note: $\mathcal{U}(\lambda_\nu)$ is an invariant subspace under $\underline{A}$, hence under all powers of $\underline{A}$, hence under all functions of $\underline{A}$.)

In a Jordan basis, $\underline{A}_\nu - \lambda_\nu = \begin{pmatrix} 0 & 1 & & & \mathbf{0} \\ & 0 & 0 & & \\ & & 0 & 1 & \\ \mathbf{0} & & & \ddots & \ddots \end{pmatrix}$. Let's concentrate on a single Jordan block, say $N = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. We find

$$N^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad N^3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = N^n \quad \text{for all } n > 2.$$

In basis-vector terms, this observation is

$$\underline{N}\vec{v}_j = \vec{v}_{j-1} \quad \text{for } j>1, \qquad \underline{N}\vec{v}_1 = \vec{0};$$

$$\underline{N}^2\vec{v}_j = \vec{v}_{j-2} \quad \text{for } j>2, \qquad \underline{N}^2\vec{v}_2 = \vec{0} = \underline{N}^2\vec{v}_1\,;$$

etc. We see that $\underline{N}$ is *nilpotent*: $\underline{N}^p = \underline{0}$ for $p \geq$ size of the Jordan block. Therefore, $\underline{A}_\nu - \lambda_\nu$ is nilpotent. It follows that the Taylor series for $f(\underline{A}_\nu)$ is a *finite sum*!

Now we can write $f(\underline{A}) = \sum_{\nu=1}^{L} \hat{f}(\underline{A}_\nu)$ (a direct-sum, or block-diagonal, operator). Let's rewrite this result in terms of projections:

Theorem. *Let*

$$\underline{A} = \sum_{\nu=1}^{L} (\lambda_\nu \underline{P}_\nu + \underline{N}_\nu)$$

be the Jordan decomposition of $\underline{A}$*. That is,* $\underline{P}_\nu$ = *projection onto* $\mathcal{U}(\lambda_\nu)$ *along the other* $\mathcal{U}(\lambda)$*'s, and* $\underline{N}_\nu \equiv \hat{\underline{A}}_\nu - \lambda_\nu \underline{P}_\nu$ *is the associated nilpotent operator on* $\mathcal{U}(\lambda_\nu)$, $\underline{A}_\nu - \lambda_\nu$, *extended as* $\underline{0}$ *to the rest of* $\mathcal{V}$*. (In general,* $\underline{N}_\nu$ *is a direct sum of elementary Jordan nilpotents.) Then*

$$f(\underline{A}) = \sum_{\nu=1}^{L} \left[f(\lambda_\nu) \underline{P}_\nu + \sum_{n=1}^{d_\nu - 1} \frac{1}{n!} f^{(n)}(\lambda_\nu)\, \underline{N}_\nu{}^n \right]. \qquad (\#)$$

(If $\underline{N}_\nu$ *is a nontrivial direct sum,* d_ν *can be replaced by the size of the largest Jordan block associated to* λ_ν*.)*

Recall that d_ν is the algebraic multiplicity of λ_ν. Note that the matrix of $f(\underline{A})$ in the Jordan basis is not necessarily in Jordan canonical form, but is block-diagonal and upper-triangular.

(#) is an *operator* equation; it holds without reference to a Jordan basis. Nevertheless, transforming to Jordan canonical form is the most obvious way to *calculate* the right-hand side of (#).

Now let's look at some applications of this theorem and of the concept of *function of an operator*:

Proof of the Hamilton–Cayley theorem

Recall that this theorem says that $\underline{A}$ satisfies its own characteristic equation, $\det(\underline{A} - \lambda) = 0$. Since Galperin and Waksman proved the Jordan theorem for us without using the HC theorem, we may use Jordan canonical form to prove HC. (Contrast Bowen & Wang.)

$$0 = \det(\underline{A} - \lambda) = (-1)^{\dim \mathcal{V}} \prod_{\nu=1}^{L} (\lambda - \lambda_\nu)^{d_\nu} \equiv f(\lambda).$$

Therefore,

$$f(\underline{A}) \equiv \pm \prod_{\nu=1}^{L} (\underline{A} - \lambda_\nu)^{d_\nu}.$$

For $\vec{v} \in \mathcal{U}(\lambda_\mu)$, we have

$$f(\underline{A})\vec{v} = \pm \prod_{\nu \neq \mu} (\underline{A} - \lambda_\nu)^{d_\nu} (\underline{A} - \lambda_\mu)^{d_\mu}\, \vec{v}.$$

But $(\underline{A} - \lambda_\mu)^{d_\mu}\, \vec{v} \equiv \underline{N}_\mu{}^{d_\mu}\, \vec{v} = \vec{0}$. Since $\mathcal{V} = \mathcal{U}(\lambda_1) \oplus \cdots \oplus \mathcal{U}(\lambda_L)$, it follows that $f(\underline{A})$ is identically $\underline{0}$, QED.

Solving systems of ordinary differential equations

References: Noble & Daniel, Sec. 10.7.
Williamson and Trotter, *Multivariable Mathematics*, Chap. 16.
Boyce and DiPrima, Chap. 7.

Example:
$$\frac{dx}{dt} = 5x + 9y$$
$$\frac{dy}{dt} = x + 3y$$

In vector notation, $\dfrac{d\vec{x}}{dt} = \underline{A}\vec{x}$. Here $\vec{x} = \vec{x}(t)$, and $\underline{A}$ is a **constant** (i.e., t-independent) linear operator from $\mathbf{R}^2$ into $\mathbf{R}^2$.

Suppose $\underline{A}$ can be diagonalized: $SAS^{-1} = D$. Define new variables by $\vec{u} \equiv S\vec{x}$. Then $d\vec{u}/dt = D\vec{u}$. In components this is of the form

$$\frac{du}{dt} = \lambda u, \qquad \frac{dv}{dt} = \kappa v,$$

where λ and κ are constants. Therefore,

$$u(t) = u(0)\, e^{\lambda t}, \qquad v(t) = v(0)\, e^{\kappa t}.$$

I.e.,

$$\vec{u}(t) = e^{tD}\, \vec{u}(0).$$

This relation is independent of basis, so

$$\vec{x}(t) = e^{t\underline{A}}\, \vec{x}(0). \qquad (*)$$

[In detail: $\vec{x}(t) = S^{-1}\vec{u}(t) = S^{-1}\, e^{tD}\, S\vec{x}(0) = e^{tS^{-1}DS}\, \vec{x}(0) = e^{tA}\, \vec{x}(0)$. It is important to remember to transform back to the $\vec{x}$ coordinates before imposing numerical initial conditions or interpreting the answer.]

The formula $(*)$ can be interpreted in two ways:

1. as shorthand for the process of diagonalizing A, finding $\vec{u}(t)$, and calculating $\vec{x}(t)$ as $S^{-1}\, \vec{u}(t)$;

2. as a direct formula for $\vec{x}(t)$, without reference to diagonalization. It's useful if we can calculate $e^{t\underline{A}}$ directly — say by power series. It's also useful for theoretical purposes when we don't need to calculate the solution explicitly.

Now suppose $\underline{A}$ can't be diagonalized:

$$SAS^{-1} = D + N = \begin{pmatrix} \lambda & 1 & 0 \\ & \ddots & \ddots \\ 0 & & \ddots \end{pmatrix} \equiv M.$$

Claim: $\vec{x}(t) = e^{t\underline{A}}\,\vec{x}(0)$ still.

Let's choose a Jordan basis and investigate this claim. The formula becomes

$$\vec{u}(t) = e^{tM}\vec{u}(0).$$

Let's evaluate this and compare it with a direct solution of the differential equation:

(1) According to our basic formula (#) for a function of an operator,

$$\begin{aligned} e^{t\underline{A}} &= \sum_{\nu=1}^{L}\left[e^{t\lambda_\nu}\underline{P}_\nu + \sum_{n=1}^{d_\nu - 1}\frac{1}{n!}\left(\frac{d}{d\lambda}\right)^n e^{t\lambda_\nu}\underline{N}_\nu{}^n\right] \\ &= \sum_{\nu=1}^{L}\left[e^{t\lambda_\nu}\underline{P}_\nu + \sum_{n=1}^{d_\nu - 1}\frac{1}{n!}\,t^n e^{t\lambda_\nu}\underline{N}_\nu{}^n\right]. \end{aligned}$$

In a Jordan basis, the matrix of $e^{t\underline{A}}$ is e^{tM}, where each $p \times p$ Jordan block of e^{tM} looks like

$$e^{t\lambda_\nu}\begin{pmatrix} 1 & t & \frac{1}{2}t^2 & \frac{1}{6}t^3 & \cdots & \frac{1}{(p-1)!}t^{p-1} \\ & 1 & t & \frac{1}{2}t^2 & & \\ & & 1 & t & \ddots & \\ & & & 1 & \ddots & \\ 0 & & & & \ddots & \end{pmatrix}$$

for some $p \le d_\nu$.

(2) We can solve the equation by elementary means if we work in the Jordan basis. For simplicity let's consider a 3×3 example,

$$M = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}; \qquad \lambda_\nu = \lambda, \qquad \vec{u} \equiv \begin{pmatrix} u(t) \\ v(t) \\ w(t) \end{pmatrix}.$$

Then $d\vec{u}/dt = M\vec{u}$ means

$$\begin{aligned} \frac{du}{dt} &= \lambda u + v, \\ \frac{dv}{dt} &= \lambda v + w, \\ \frac{dw}{dt} &= \lambda w. \end{aligned}$$

[Note: u, v, w are *not* basis vectors, but rather components of the vector-valued function

$$\vec{u} = u\,\vec{e}_1 + v\,\vec{e}_2 + w\,\vec{e}_3\,,$$

where the basis vectors satisfy the Jordan-chain relations

$$M\vec{e}_3 = \lambda\vec{e}_3 + \vec{e}_2\,, \qquad M\vec{e}_2 = \lambda\vec{e}_2 + \vec{e}_1\,, \qquad M\vec{e}_1 = \lambda\vec{e}_1\,.$$

As usual, the matrix acting on the coordinates is the transpose of that acting on the basis vectors.] We can easily solve this system from the bottom up:

$$w(t) = w(0)e^{\lambda t}\,;$$

$$\frac{dv}{dt} = \lambda v + w(0)e^{\lambda t},$$
$$v(t) = v(0)e^{\lambda t} + w(0)te^{\lambda t}\,;$$

$$\frac{du}{dt} = \lambda u + v(0)e^{\lambda t} + w(0)te^{\lambda t},$$
$$u(t) = u(0)e^{\lambda t} + v(0)te^{\lambda t} + \frac{1}{2}w(0)t^2e^{\lambda t}.$$

Thus

$$\begin{pmatrix} u \\ w \\ w \end{pmatrix} = e^{\lambda t}\begin{pmatrix} 1 & t & \frac{1}{2}t^2 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} u(0) \\ v(0) \\ w(0) \end{pmatrix}$$
$$\equiv e^{tM}\vec{u}(0) = \vec{u}(t),$$

in agreement with (1). Clearly the argument is general, although our presentation was for an example.

The concept of the exponential operator as solution operator generalizes to partial differential equations: Consider

$$\frac{\partial u}{\partial t} = -\underline{A}\,u, \qquad u = u(t,x), \qquad u(0,x) = f(x) \text{ given},$$

where $\underline{A}\colon \mathcal{L}^2(\mathbf{R}^m) \to \mathcal{L}^2(\mathbf{R}^m)$ is a positive, self-adjoint operator (e.g., $\underline{A} = -\nabla^2$). The solution can be written

$$u(t,x) = [e^{-t\underline{A}}\,f](x).$$

Rather than use this formula to obtain an explicit numerical solution, the usual application is in the reverse direction: Direct study of $e^{-t\underline{A}}$ by PDE methods gives useful information about $\underline{A}$ and its eigenfunctions.

Another generalization (to second-order ODEs) is indicated in one of the homework problems. It involves functions other than the exponential.