Newton

and the

Characteristic Polynomial of a Matrix

Nicholas Wheeler
December 2019

Introduction. My title is, of course, anachronistic: matrices and their characteristic polynomials did not enter into the thought of mathematicians until a full century after Newton's death. Both, however, derive from a concept—that of the "determinant"—which, we are informed,[1] was implicit in the efforts of Chinese mathematicians to solve systems of simultaneous linear equations already by the 3rd century BC. Two millennia elapsed before determinants began to appear in the work—related again to that same problem—of Western mathematicians (Cardano at the end of the 16th century, Leibniz in the 17th). Determinants figure in the 18th century work of Vandermonde, Laplace and Lagrange, but it was not until the 19th century that the subject came to be studied in a systematic way, and to flourish: on 30 November 1812 Jacques Binet and Arthur Cayley gave independent accounts of the basic properties of determinants (it was Cayley who gave them their name), whereupon the subject was taken up by Hamilton, Grassmann, Sylvester, Cramer, Gauss and many others.

The theory of matrices came later, born of the theory of determinants. Only in 1856 did Cayley describe matrix multiplication and inversion, and take the important step of denoting such objects with a single symbol. He wrote "There would be many things to say about this theory of matrices which should, it seems to me, precede the theory of determinants." Only as those things began to be said did characteristic polynomials and their roots, together with much else—including, as will emerge, the ghost of Newton—enter the picture.

---

[1] See the Wikipedia articles "Determinant" and "Linear algebra."

Foreshadowing things to come. Let $\mathbb{A}$ be a 2×2 matrix:

$$\mathbb{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

Then

$$\begin{aligned}
\det\mathbb{A} &= a_{11}a_{22} - a_{12}a_{21} && (1)\\
\mathrm{tr}\mathbb{A} &= a_{11} + a_{22} && (2.1)\\
\mathrm{tr}\mathbb{A}^2 &= a_{11}^2 + 2a_{11}a_{22} + a_{22}^2 = (a_{11}+a_{22})^2 && (2.2)\\
\mathrm{tr}\mathbb{A}^3 &= a_{11}^3 + 3(a_{11}+a_{22})a_{12}a_{21} + a_{22}^3 \neq (a_{11}+a_{22})^3
\end{aligned}$$

from which it follows in particular that

$$\det\mathbb{A} = \tfrac{1}{2}\big[(\mathrm{tr}\mathbb{A})^2 - \mathrm{tr}\mathbb{A}^2\big] \tag{3}$$

The characteristic polynomial is

$$\begin{aligned}
\det(\lambda\mathbb{I} - \mathbb{A}) &= \lambda^2 - (a_{11}+a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21})\\
&= \lambda^2 - \lambda\,\mathrm{tr}\mathbb{A} + \det\mathbb{A}\\
&= \lambda^2 - \lambda\,\mathrm{tr}\mathbb{A} + \tfrac{1}{2}\big[(\mathrm{tr}\mathbb{A})^2 - \mathrm{tr}\mathbb{A}^2\big] && (4)
\end{aligned}$$

of which the roots (which we would find it difficult or impossible to write down in higher-dimensional cases) are

$$\left.\begin{aligned}
\lambda_1 &= \tfrac{1}{2}\Big[(a_{11}+a_{22}) + \sqrt{(a_{11}-a_{22})^2 + 4a_{12}a_{21}}\Big]\\
\lambda_2 &= \tfrac{1}{2}\Big[(a_{11}+a_{22}) - \sqrt{(a_{11}-a_{22})^2 + 4a_{12}a_{21}}\Big]
\end{aligned}\right\} \tag{5}$$

These entail

$$\begin{aligned}
\lambda_1 + \lambda_2 &= \mathrm{tr}\mathbb{A} && (6.1)\\
\lambda_1\lambda_2 &= \det\mathbb{A} && (6.2)
\end{aligned}$$

which also follow directly from comparison of (4) with

$$\begin{aligned}
\det(\lambda\mathbb{I} - \mathbb{A}) &= (\lambda - \lambda_1)(\lambda - \lambda_2) && (7)\\
&= \lambda^2 - \lambda(\lambda_1 + \lambda_2) + \lambda_1\lambda_2
\end{aligned}$$

In λ-language (3) has become

$$\lambda_1\lambda_2 = \tfrac{1}{2}\big[(\lambda_1+\lambda_2)^2 - (\lambda_1^2 + \lambda_2^2)\big] \tag{8}$$

Newton—apparently unaware that he was duplicating/extending results that were known to Albert Girard (1595–1632) already by 1629—in about 1666 pioneered the theory of symmetric polynomials, and it is because $\{\lambda_1, \lambda_2\}$ enter symmetrically into (6–8) that Newton enters the picture. My objective in these pages will be to indicate how **Newton's identitiies** illuminate the higher-dimensional analogs of (6–8).

Plan of attack: Tracewise construction of determinants. If $\mathbb{A}$ is $n\times n$ then the definition

$$\det\mathbb{A} = \sum_{\mathcal{P}}(-)^{\mathcal{P}} a_{1k_1}a_{2k_2}a_{3k_3}\cdots a_{nk_n}$$

—here $\sum_{\mathcal{P}}$ signifies summation over all $n!$ of the permutations

$$\mathcal{P} = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ k_1 & k_2 & k_3 & \dots & k_n \end{pmatrix}$$

and $(-)^{\mathcal{P}}$ is the signature (parity) of $\mathcal{P}$—shows the determinant $\det\mathbb{A}$ to be a profoundly *antisymmetric* function of the elements a_{ij} of $\mathbb{A}$. On the other hand,

$$\det\mathbb{A} = \lambda_1\lambda_2\cdots\lambda_n$$

shows $\det\mathbb{A}$ to be a profoundly *symmetric* function of the eigenvalues λ_i of $\mathbb{A}$. A similarly constrasting distinction pertains to these descriptions

$$\det(\lambda\mathbb{I}-\mathbb{A}) \quad \textit{vs.} \quad \prod_{k=1}^{n}(\lambda-\lambda_k)$$

of the characteristic polynomial.

To bring these consructions into harmony my strategy will be (i) to write

$$\det(\lambda\mathbb{I}-\mathbb{A}) = \lambda^n + \sum_{m=1}^{n} D_m\lambda^{n-m}$$

(ii) to develop D_m as a function of $\{T_1, T_2, \dots, T_m\} : T_k \equiv \mathrm{tr}\,\mathbb{A}^k$, and (iii) to use

$$T_k \equiv \mathrm{tr}\,\mathbb{A}^k = \lambda_1^k + \lambda_2^k + \cdots + \lambda_n^k$$

Development of the coefficients D_m can be accomplished by an argument I devised in 1958 (thinking—mistakenly—that I was the first to venture down that trail[2]), and have on several occasions refined to serve a variety of applications. Most readily accessible is some material I wrote to be of service to Richard Crandall,[3] from which I will be content here simply to quote. The

[2] The first appears to have been Urbain Le Verrier (1811–1877), whose calculations led to the discovery of Neptune (1846) and to recognition of the anomalous precession of the perihelion of Mercury (1859). It was upon a foundation laid by Le Verrier that Dmitry Faddeev (1907–1989) erected the "Le Verrier-Faddeev characteristic polynomial algorithm."

[3] "A Mathematical Note: Algorithm for the efficient evaluation of the trace of the inverse of a matrix," (December, 1996). That paper provides refrences to several previous discussions of this material.

argument proceeds

$$
\begin{aligned}
\det(\lambda\mathbb{I}-\mathbb{A}) = \lambda^n \det(\mathbb{I}-x\,\mathbb{A}) \quad &: \quad x \equiv 1/\lambda\\
\det(\mathbb{I}-x\,\mathbb{A}) &= \exp\big\{\mathrm{tr}\log(\mathbb{I}-x\,\mathbb{A})\big\}\\
&= \exp\big\{-T_1x-\tfrac{1}{2}T_2x^2-\tfrac{1}{3}T_3x^3-\tfrac{1}{4}T_4x^4-\cdots\big\}\\
&= 1-\tfrac{1}{1!}\big[T_1\big]x\\
&\qquad+\tfrac{1}{2!}\big[T_1^2-T_2\big]x^2\\
&\qquad-\tfrac{1}{3!}\big[T_1^3-3T_1T_2+2T_3\big]x^3\\
&\qquad+\tfrac{1}{4!}\big[T_1^4-6T_1^2T_2+3T_2^2+8T_1T_3-6T_4\big]x^4\\
&\qquad\vdots\\
&\qquad(-)^n\big[\text{stuff}\big]x^n
\end{aligned}
$$

giving

$$
\begin{aligned}
\det(\lambda\mathbb{I}-\mathbb{A}) &= \lambda^n-\tfrac{1}{1!}\big[T_1\big]\lambda^{n-1}\\
&\qquad+\tfrac{1}{2!}\big[T_1^2-T_2\big]\lambda^{n-2}\\
&\qquad-\tfrac{1}{3!}\big[T_1^3-3T_1T_2+2T_3\big]\lambda^{n-3}\\
&\qquad+\tfrac{1}{4!}\big[T_1^4-6T_1^2T_2+3T_2^2+8T_1T_3-6T_4\big]\lambda^{n-4}\\
&\qquad\vdots\\
&\qquad(-)^n\big[\text{stuff}\big]\lambda^0\\
&= \lambda^n+\sum_{m=1}^n D_m\lambda^{n-m}
\end{aligned}
$$

where the coefficients can be described

$$
\begin{aligned}
D_m &= (-)^m\tfrac{1}{m!}\begin{vmatrix} T_1 & T_2 & T_3 & T_4 & \dots & T_m\\ 1 & T_1 & T_2 & T_3 & \dots & T_{m-1}\\ 0 & 2 & T_1 & T_2 & \dots & T_{m-2}\\ 0 & 0 & 3 & T_1 & \dots & T_{m-3}\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & 0 & \dots & T_1 \end{vmatrix} \quad &: \quad m=1,2,\dots,n\\
&= 0 &: \quad m>n
\end{aligned}
\tag{9.1}
$$

It is obvious that in the n-dimensional case $D_n=(-)^n\big[\text{stuff}\big]=\det\mathbb{A}$, while $D_{m>n}=0$ is easily seen to follow from the Cayley-Hamilton theorem:

$$
\mathbb{A}^n+\sum_{m=1}^n D_m\mathbb{A}^{n-m}=\mathbb{O}
$$

In the case $n=2$ we recover (4).

The eigenvalue representation. The results developed on the preceding page are formulated in tems of the elements a_{ij} of $\mathbb{A}$. But if $\mathbb{A}$ is diagonal, or can be brought to diagonal form by a similarity transformation $\mathbb{S}^{-1}\mathbb{A}\mathbb{S}$, it follows from this general property $\mathrm{tr}(\mathbb{A}\mathbb{B}) = \mathrm{tr}\mathbb{B}\mathbb{A})$ of the trace that (as was remarked already on page 3)

$$T_k \equiv \mathrm{tr}\,\mathbb{A}^k = \lambda_1^k + \lambda_2^k + \cdots + \lambda_n^k \tag{9.2}$$

the validity of which is actually unrestricted.[4] On the other hand, we have

$$\begin{aligned}
\det(\lambda\mathbb{I} - \mathbb{A}) = \prod_{k=1}^{n}(\lambda - \lambda_k) = \lambda^n &- \lambda^{n-1}\sum_{1\leqslant k\leqslant n}\lambda_k \\
&+ \lambda^{n-2}\sum_{1\leqslant k_1<k_2\leqslant n}\lambda_{k_1}\lambda_{k_2} \\
&- \lambda^{n-3}\sum_{1\leqslant k_1<k_2<k_3\leqslant n}\lambda_{k_1}\lambda_{k_2}\lambda_{k_3} \\
&\;\;\vdots \\
&(-)^n\lambda^0\cdot\lambda_1\lambda_2\cdots\lambda_n
\end{aligned} \tag{10}$$

where (since the λ's commute)

$$\begin{aligned}
\sum_{1\leqslant k\leqslant n}\lambda_k &\text{ is a symmetric sum of } \tbinom{n}{1}\ 1^{\text{st}}\text{-order terms} \\
\sum_{1\leqslant k_1<k_2\leqslant n}\lambda_{k_1}\lambda_{k_2} &\text{ is a symmetric sum of } \tbinom{n}{2}\ 2^{\text{nd}}\text{-order terms} \\
\sum_{1\leqslant k_1<k_2<k_3\leqslant n}\lambda_{k_1}\lambda_{k_2}\lambda_{k_3} &\text{ is a symmetric sum of } \tbinom{n}{3}\ 3^{\text{rd}}\text{-order terms} \\
&\;\;\vdots \\
\lambda_1\lambda_2\cdots\lambda_n &\text{ is a symmetric sum of } \tbinom{n}{n}\ n^{\text{th}}\text{-order terms}
\end{aligned}$$

In (9) and (10) we have *two* alternative eigenvalue-based descriptions of the coefficients D_m that enter into the construction of the characteristic polynomial of the $n\times n$ matrix $\mathbb{A}$. We turn now to discussion of their evident equivalence.

Let $\mathcal{X}$ signify the n-element set of variables $\{x_1, x_2, \ldots, x_n\}$. *Elementary symmetric polynomials* are multinomials of the form encountered in (10):

$$\begin{aligned}
e_0(\mathcal{X}) &= 1 \\
e_1(\mathcal{X}) &= \textstyle\sum\lambda_k &&: \quad 1\leqslant k\leqslant n \\
e_{m\leqslant n}(\mathcal{X}) &= \textstyle\sum\lambda_{k_1}\lambda_{k_2}\cdots\lambda_{k_m} &&: \quad 1\leqslant k_1<k_2<\cdots<k_m\leqslant n \\
e_{m>n}(\mathcal{X}) &= 0
\end{aligned}$$

[4] A similar remark pertains to the identity $\det\mathbb{A} = e^{\mathrm{tr}(\log\mathbb{A})}$ which provided our point of departure.

At (9.2) we encountered symmetric polynomials of the form

$$p_k(\mathfrak{X}) = x_1^k + x_2^k + \cdots + x_n^k \quad : \quad k^{\text{th}} \textit{ power sum}$$

Symmetric polynomials in the variables $\mathfrak{X}$ exist in infinite variety.[5] The set of such objects is closed under addition and multiplication, so is a ring, $\mathcal{R}_{\mathfrak{X}}$. The elementary symmetric functions $e_i(\mathfrak{X}) : i = 0, 1, 2, \ldots, n$ ascquire special importance from the **fundamental theorem of symmetric polynomials**, which asserts[6] that every element $q(\mathfrak{X}) \in \mathcal{R}_{\mathfrak{X}}$ admits of unique formulation as a polynomial combination of elementary polynomials:

$$q(\mathfrak{X}) = P(e_0(\mathfrak{X}), e_1(\mathfrak{X}), \ldots, e_n(\mathfrak{X}))$$

Contact with Newton. Let $\mathcal{L}$ signify the set of variables $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$. Reading from (10) and (9.1), which provide alternative descriptions of the D_k in the characteristic polynomial of the n-dimensional matrix $\mathbb{A}$, we are in position now to write (for $m = 1, 2, \ldots n$)

$$e_m(\mathcal{L}) = \tfrac{1}{m!} \begin{vmatrix} p_1(\mathcal{L}) & p_2(\mathcal{L}) & p_3(\mathcal{L}) & p_4(\mathcal{L} & \ldots & p_m(\mathcal{L}) \\ 1 & p_1(\mathcal{L}) & p_2(\mathcal{L}) & p_13(\mathcal{L}) & \ldots & p_{m-1}(\mathcal{L}) \\ 0 & 2 & p_1(\mathcal{L}) & p_2(\mathcal{L}) & \ldots & p_{m-2}(\mathcal{L}) \\ 0 & 0 & 3 & p_1(\mathcal{L}) & \ldots & p_{m-3}(\mathcal{L}) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & p_1(\mathcal{L}) \end{vmatrix} \tag{11.1}$$

and $e_{m>n}(\mathcal{L}) = 0$. Which, when spelled out (if we omit the arguments), read

$$\left.\begin{aligned} e_1 &= p_1 \\ e_2 &= \tfrac{1}{2!}[p_1^2 - p_2] \\ e_3 &= \tfrac{1}{3!}[p_1^3 - 3p_1p_2 + 2p_3] \\ e_4 &= \tfrac{1}{4!}[p_1^4 - 6p_1^2p_2 + 3p_2^2 + 8p_1p_3 - 6p_4] \\ &\ \vdots \\ e_m &= 0 \quad : \quad m > n \end{aligned}\right\} \tag{11.2}$$

Equations (11.2) are precisely the **Newton identities** that describe the elementary symmetric polynomials as polynomials in power sums, and so inform us that if—as the fundamental theorem asserts–the polynomials $e_m(\mathcal{L})$ provide an algebraic basis in $\mathcal{R}_{\mathcal{L}}$ then so also do the polynomials $p_k(\mathcal{L})$. The equations (11.2) are

[5] Take an arbitrary polynomial in variables $\{x_1, x_2, \ldots, x_{m\leqslant n}\}$, hit it with $\sum_{\mathcal{P}}$, where $\mathcal{P}$ is a permutation of the elements of $\mathfrak{X} \equiv \{x_1, x_2, \ldots, x_n\}$.

[6] See the Wikipedia article "Fundamental theorem of symmetric polynomials."

equations (11.2) are usually obtained recursivly from

$$\left.\begin{aligned} e_1 &= p_1 \\ 2e_2 &= e_1 p_1 - p_2 \\ 3e_3 &= e_2 p_1 - e_1 p_2 + p_3 \\ 4e_4 &= e_3 p_1 - e_2 p_2 + e_1 p_3 - p_4 \\ &\vdots \end{aligned}\right\} \tag{12.1}$$

Recursive inversion of (11.2) or of (12.1) gives

$$\left.\begin{aligned} p_1 &= e_1 \\ p_2 &= e_1^2 - 2e_2 \\ p_3 &= e_1^3 - 3e_1 e_2 + 3e_3 \\ p_4 &= e_1^4 - 4e_1^2 e_2 + 4e_1 e_3 + 2e_2^2 - 4e_4 \\ &\vdots \end{aligned}\right\} \tag{12.2}$$

These are the Newton identities that had been obtained already in 1629 by Albert Girard.

Working from $\det(\mathbb{I} - x\,\mathbb{A}) = \exp\big\{\operatorname{tr}\log(\mathbb{I} - x\,\mathbb{A})\big\}$, we on page 4 obtained a result that in terms of the symmetric functions $e_m(\mathcal{L})$ and $p_k(\mathcal{L})$ can be written

$$\exp\big\{ - p_1 x - \tfrac{1}{2} p_2 x^2 - \tfrac{1}{3} p_3 x^3 - \cdots \big\} = 1 - \tfrac{1}{1!} e_1 x + \tfrac{1}{2!} e_2 x^2 - \tfrac{1}{3!} e_3 x^3 + \cdots$$

or

$$\sum_{m=0}^{\infty} e_m x^m = \exp\Big\{ \sum_{k=1}^{\infty} (-)^{k+1} \tfrac{1}{k}\, p_k x^k \Big\}$$

which generates the identities (11.2). Similarly, we have

$$\sum_{k=1}^{\infty} (-)^{k+1} \tfrac{1}{k}\, p_k x^k = \log\Big\{ \sum_{m=0}^{\infty} e_m x^m \Big\}$$

which generates (12.2).

Concluding remarks. Symmetric polynomials have been studied for nearly 400 years, so there is little fresh to be said about them, and nothing here can be said to be "fresh" except, perhaps, for my approach—from linear algebra, *via* the beautiful but nameless identity $\det(\mathbb{I} - x\,\mathbb{A}) = \exp\big\{\operatorname{tr}\log(\mathbb{I} - x\,\mathbb{A})\big\}$.[7]

These pages can be dismissed as the sentimental indulgence of an old man, fruit of the confluence of three events of some importance to me in the span

[7] For a good survey of the subject, see, for example, the Wikipedia article "Newton's identities."

of my mathematical career. In chronological order:

In the spring of 1953, I was a sophomore student in the last class taught by Robert Rosenbaum (1915–2017) before he left Reed College to join the Wesleyan faculty. The subject was projective geometry, the principal texts were *Introduction to Higher Geometry* by William Graustein (1888–1941), *Projective and Analytical Geometry* by John Todd (1908–1994), supplemented by Bob's own mimeographed notes, still damp and smelling of alcohol when they were distributed at the beginning of the hour. I was asked to present a survey of the basics of the theory of symmetric polynomials—a subject for which I developed an affection, but to which I never (until now) had occasion to return.

In 1958, while working on my doctoral dissertation, I encountered need to describe the n^{th} derivative of a composite function $F(x) = \Phi(f(x))$. While trying to track down a relevant reference in the Brandeis University library I happened by lucky accident upon Advanced Problem No. 4782 which had been submitted by V.F. Ivanoff to American Mathematical Monthly (**65**, 212 (1958)). The problem had been solved by Francesco Faà di Bruno (1825–1888: classmate of Hermite, beatified in 1988). Ivanoff asked for demonstrations (of which I supplied several) that di Bruno's result can be formulated

$$F^{(n)}(x) = \begin{vmatrix} f'D & f''D & f'''D & f''''D & \dots & f^{(n)}D \\ -1 & f'D & 2f''D & 3f'''D & \dots & \binom{n-1}{1}f^{(n-1)}D \\ 0 & -1 & f'D & 3f''D & \dots & \binom{n-1}{2}f^{(n-2)}D \\ 0 & 0 & -1 & f'D & \dots & \binom{n-1}{3}f^{(n-3)}D \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & f'D \end{vmatrix} \Phi(f) \tag{13}$$

where $D^k\Phi(f) \equiv \big(\frac{d}{df}\big)^k\Phi(f)$. Thus did ramifications[8] of (13)—determinants of that triangular design, adapted to many diverse applications—come to assume—as here—a recurrent place in my work.

Finally, symmetric polynomials were recalled to mind by the expression

$$\prod_{m=1, m\neq k}^{n} (\lambda - \lambda_m)$$

that figures prominently in my most recent work.[9] The present essay falls again—here as there—under the head "The secret lives of eigenvalues"... as a quanum physicist might argue everything in the physical world does.

[8] See "Some applicatioms of an elegant formula due to I. F. Ivanoff," notes for a seminar presented on 28 May 1969 to the Applied Math Club at Portland State University (COLLECTED SEMINARS 1963–1970).

[9] "Eigenvalues as building bricks," (December 2019).