EXTRAPOLATED INTERPOLATION THEORY

Nicholas Wheeler, Reed College Physics Department
April 1997

Introduction. Given a "seed" function $f(x)$ we might—subject to appropriate restrictions—employ operations of the ordinary calculus to construct

$$\cdots \leftarrow \iiint f \leftarrow \iint f \leftarrow \int f \leftarrow \boldsymbol{f} \rightarrow f' \rightarrow f'' \rightarrow f''' \rightarrow \cdots$$

and seek to "interpolate" between the points thus marked out in function space. The result would be a "fractional calculus." It was with that admittedly vague (and, as I discovered, not terribly original[1]) thought in mind that, prior to a recent excursion into the fractional calculus,[2] I undertook to review—from a somewhat eccentric point of view—some of the most rudimentary elements of interpolation/extrapolation theory. Here I record the results of that brief effort.

An element of ambiguity attaches characteristically and unavoidably to all interpolation/extrapolation schemes, which fall therefore within the general ruberic of the elaborately worked-out subject called "approximation theory." One looks in all cases for the scheme that pertains "most naturally" to the problem at hand, in the sense of being on the one hand "simplest," and on the other hand "most empowering." In a broad class of typical applications the objective is to construct a best-fit analytical description of observational data. That class of applications is *not* of present interest to me. I draw my

[1] Leonard Euler had expressed a similar thought when, in 1730, he wrote

> *"Concerning transcendental progressions whose terms cannot be given algebraically: when n is a positive integer, the ratio $d^n f/dx^n$ can always be expressed algebraically. Now it is asked: what kind of ratio can be made if n be a fraction? . . . the matter may be expedited with the help of the interpolation of series, as explained earlier in this dissertation."*

What Euler might have had specifically in mind by the phrase "interpolation of series" I can, however, only guess. . . and will.

[2] See "Construction & physical application of the fractional calculus" (1997).

inspiration instead from certain relatively more formal/theoretical applications of ideas borrowed from interpolation theory—applications to (for example) the description and direct geometrical interpretation of the elements $\mathbb{R}$ of the n-dimensional rotation group $O(n)$, to certain counting problems that arise in connection with the theory of hyperspherical harmonics. . . and to problems suggested by the fractional calculus. I begin somewhat obliquely:

1. Triangular and other figurate numbers. The triangular numbers $\Delta(n)$ arise familiarly from the following sequence of constructions

$$\bullet \qquad\qquad \begin{array}{ll} \bullet & \\ \bullet & \bullet \end{array} \qquad\qquad \begin{array}{lll} \bullet & & \\ \bullet & \bullet & \\ \bullet & \bullet & \bullet \end{array} \qquad\qquad \text{etc.}$$

They occur in connection with ennumerative problems of many types—the theory of (anti)symmetric matrices provides a typical instance

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ * & a_{22} & a_{23} & \cdots & a_{2n} \\ * & * & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ * & * & * & \cdots & a_{nn} \end{pmatrix}$$

—and are in themselves almost (but not quite) too simple to support theoretical statements of any interest.[3] The triangular numbers do, however, support a rich population of natural generalizations, and some (at least recreational) interest does attach to the relationships which come into view when that population is regarded as a whole.[4] For example: the triangular numbers $\{1, 3, 6, 10, \ldots\}$

[3] On another occasion I hope to discuss physical implications of the curious (and by no means elementary) fact that

$$2^p - 1 \text{ is triangular only in the cases } p = 1, 2, 4 \text{ and } 12$$

The statement $2^p - 1 = \frac{1}{2}n(n+1)$ can be formulated $2^{p+3} - 7 = (2n+1)^2$. On pp. 322–334 of the *Collected Papers of Srinivasa Ramanujan* (edited by G. H. Hardy *et al* (1927)) one finds a list of "questions and solutions submitted by Ramanujan to the Journal of the Indian Mathematical Society." Question 464 reads "$2^p - 7$ is a perfect square for the values 3, 4, 5, 7, 15 of p. Find other values." This has traditionally been interpreted as a conjecture that there exist *no* other values. The conjecture was proved correct by T. Nagell in Norsk Matematisk Tidsskrift **30**, 62 (1948), and (independently) by T. Skolem, S. Chowla & D. Lewis, Proc. Amer. Math. Soc. **10**, 663 (1959). D. Lewis, Pacific Journal of Math. **11**, 1063 (1961) shows Ramanujan's conjecture to comprise only the simplest instance of a general class of related problems.

[4] Tom Apostol, in the Historical Introduction to his *Introduction to Analytic Number Theory* (1976), remarks that figurate numbers were of special interest already to the Pythagoreans because they served to link numbers with geometry.

arise by partial summation from the sequence

$$\mathbf{1},\mathbf{2},\mathbf{3},\mathbf{4},\mathbf{5},\mathbf{6},\mathbf{7},\mathbf{8},\mathbf{9},\mathbf{10},\mathbf{11},\mathbf{12},\mathbf{13},\mathbf{14},\mathbf{15},\mathbf{16},\ldots$$

Partial summation on the related arithmetic sequence

$$\mathbf{1},*,\mathbf{3},*,\mathbf{5},*,\mathbf{7},*,\mathbf{9},\ *\ ,\mathbf{11},\ *\ ,\mathbf{13},\ *\ ,\mathbf{15},\ *\ ,\ldots$$

gives rise on the other hand to the sequence $\{1,4,9,16,\ldots\}$ of "square numbers"

$$\begin{array}{c} \bullet \\ \\ \\ \end{array} \qquad \begin{array}{cc} \bullet & \circ \\ \circ & \circ \\ \\ \end{array} \qquad \begin{array}{ccc} \bullet & \bullet & \circ \\ \bullet & \bullet & \circ \\ \circ & \circ & \circ \end{array} \qquad \begin{array}{c} \text{etc.} \\ \\ \\ \end{array}$$

while summation on the arithmetic sequence

$$\mathbf{1},*,*,\mathbf{4},*,*,\mathbf{7},*,*,\mathbf{10},\ *\ ,\ *\ ,\mathbf{13},\ *\ ,\ *\ ,\mathbf{16},\ldots$$

generates the sequence $\{1,5,12,22,\ldots\}$ of "pentagonal numbers," summation on the sequence

$$\mathbf{1},*,*,*,\mathbf{5},*,*,*,\ \mathbf{9},\ *\ ,\ *\ ,\ *\ ,\mathbf{13},\ *\ ,\ *\ ,\ *\ ,\ldots$$

generates the sequence $\{1,6,15,28,\ldots\}$ of "hexagonal numbers," etc. More generally, partial summation on the arithmetic sequence

$$1, 1+a, 1+2a, 1+3a,\ldots$$

gives rise to the sequence $\{1, 2+a, 3+3a, 4+6a,\ldots\}$ of "polyhedral numbers of order $N = a+2$." The notations $\Delta(n)$, $\Box(n)$, ... do not readily generalize, so we agree to write

$$\begin{aligned} P(n;3) &= n^{\text{th}} \text{ triangular number} \equiv \Delta(n) \\ P(n;4) &= n^{\text{th}} \text{ square number} \\ &\ \,\vdots \\ P(n;N) &= n^{\text{th}}\text{polygonal number of order } N \end{aligned}$$

By "stacking" triangles of graded size we obtain the "tetrahedral numbers"

$$T(n;3) \equiv \sum_{k=1}^{n} \Delta(k)$$

and if we pass into the 4^{th}-dimension and proceed to "stack tetrahedra" we obtain the "hypertetrahedral numbers"

$$T(n;4) \equiv \sum_{k=1}^{n} T(k;3)$$

The procedure just outlined can obviously be extended to arbitrary dimension D. Polyhedral numbers of more general design ($N > 3$) support a similar program of dimensional extension; the associated figures are "hyperpyramids" with polyonal bases. . . and are, for my purposes, relatively uninteresting.

It was obvious already to the infant Gauss that

$$\begin{aligned}2\Delta(n) = {}& \quad 1 + \qquad 2\ + \qquad 3\ + \cdots + n \\ & + n + (n-1) + (n-2) + \cdots + 1 \\ = {}& n(n+1)\end{aligned}$$

which gives the familiar result

$$\Delta(n) = \tfrac{1}{2}n(n+1) \tag{1}$$

So also does the construction

$$\begin{array}{cccc} \circ & \circ & \circ & \circ \\ \bullet & \circ & \circ & \circ \\ \bullet & \bullet & \circ & \circ \\ \bullet & \bullet & \bullet & \circ \\ \bullet & \bullet & \bullet & \bullet \end{array}$$

And from the structure of the sequence that generates $P(n;N)$ it is clear that

$$P(n;N) = n + a\Delta(n-1)$$

From this it follows in particular that

$$\begin{aligned}P(n;3) &= n + 1\cdot\tfrac{1}{2}(n-1)n = \tfrac{1}{2}n(1n+1) = \tfrac{1}{2}n(n+1)\\ P(n;4) &= n + 2\cdot\tfrac{1}{2}(n-1)n = \tfrac{1}{2}n(2n+0) = n^2\\ P(n;5) &= n + 3\cdot\tfrac{1}{2}(n-1)n = \tfrac{1}{2}n(3n-1)\\ P(n;6) &= n + 4\cdot\tfrac{1}{2}(n-1)n = \tfrac{1}{2}n(4n-2)\\ P(n;7) &= n + 5\cdot\tfrac{1}{2}(n-1)n = \tfrac{1}{2}n(5n-3)\\ &\ \vdots\end{aligned}$$

Looking now in greater detail to the "tetrahedral numbers," we have

$$\begin{aligned}T(n;1) &\equiv n \quad : \quad \text{a formal convenience}\\ T(n;2) &\equiv \sum_{k=1}^{n} T(k;1) \equiv \Delta(n) = \tfrac{1}{2}n(n+1) = \binom{n+1}{2}\\ T(n;3) &\equiv \sum_{k=1}^{n} T(k;2) = \sum_{k=1}^{n} \tfrac{1}{2}k(k+1)\\ &\qquad\qquad\qquad = \tfrac{1}{6}n(1+n)(2+n)\\ &\qquad\qquad\qquad = \tfrac{1}{3}n + \tfrac{1}{2}n^2 + \tfrac{1}{6}n^3\\ &\qquad\qquad\qquad \sim \tfrac{1}{3!}n^3 \quad \text{for } n \text{ large}\end{aligned}$$

$$
\begin{aligned}
T(n,4) \equiv \sum_{k=0}^{n} T(n,3) &= \tfrac{1}{24}n(1+n)(2+n)(3+n) \\
&= \tfrac{1}{4}n + \tfrac{11}{24}n^2 + \tfrac{1}{4}n^3 + \tfrac{1}{24}n^4 \\
&\sim \tfrac{1}{4!}n^4 \quad \text{for } n \text{ large} \\
T(n,5) \equiv \sum_{k=0}^{n} T(n,4) &= \tfrac{1}{120}n(1+n)(2+n)(3+n)(4+n) \\
&= \tfrac{1}{5}n + \tfrac{5}{12}n^2 + \tfrac{7}{24}n^3 + \tfrac{1}{12}n^4 + \tfrac{1}{120}n^5 \\
&\sim \tfrac{1}{5!}n^5 \quad \text{for } n \text{ large} \\
&\;\vdots \\
T(n,D) \equiv \sum_{k=0}^{n} T(k,D-1) &= \tfrac{1}{D!}n(1+n)(2+n)(3+n)\cdots([D-1]+n) \\
&= \tfrac{(D+n-1)!}{D!(n-1)!} = \binom{D+n-1}{D} \\
&\sim \tfrac{1}{D!}n^D \quad \text{for } n \text{ large}
\end{aligned}
$$

The "Pochhammer polynomials" are defined[5] $(x)_n \equiv x(x+1)(x+2)\cdots(x+n-1)$ and, in view of the preceding results, permit us to write

$$T(n,D) = \tfrac{1}{D!}(n)_D$$

The same information can be expressed also recursively:

$$
\begin{aligned}
T(n,2) &= T(n,1)\cdot\frac{1+n}{2} \\
T(n,3) &= T(n,2)\cdot\frac{2+n}{3} = T(n,1)\cdot\frac{1+n}{2}\cdot\frac{2+n}{3} \\
T(n,4) &= T(n,3)\cdot\frac{3+n}{4} = T(n,1)\cdot\frac{1+n}{2}\cdot\frac{2+n}{3}\cdot\frac{3+n}{4} \\
&\;\vdots
\end{aligned}
$$

If, instead of stacking triangles, we stack squares we are led (by dimensional generalization; i.e, by successively stacking cubes, hypercubes, etc.) to the "pyramidal numbers" of ascending order:

$$
\begin{aligned}
S(n;2) &\equiv \sum_{k=1}^{n} k^2 \\
S(n;3) &\equiv \sum_{k=1}^{n} k^3 \\
&\;\vdots
\end{aligned}
$$

[5] See J. Spanier & K. Oldham, *Atlas of Functions* (1987), Chapter 18.

We are informed by *Mathematica* that

$$\begin{aligned} S(n,0) &= n \\ \Delta(n) = S(n,1) &= \tfrac{1}{2}n(1+n) \\ S(n,2) &= \tfrac{1}{6}n(1+n)(1+2n) \\ S(n,3) &= \tfrac{1}{4}n^2(1+n)^2 \\ &= [S(n,1)]^2 \\ S(n,4) &= \tfrac{1}{5}S(n,2)\cdot\text{polynomial of degree 2} \\ S(n,5) &= \tfrac{1}{3}S(n,3)\cdot\text{polynomial of degree 2} \\ S(n,6) &= \tfrac{1}{7}S(n,2)\cdot\text{polynomial of degree 4} \\ S(n,7) &= \tfrac{1}{6}S(n,3)\cdot\text{polynomial of degree 4} \\ &\vdots \end{aligned}$$

I have here surpressed certain details in order to highlight a factorization pattern which was striking news to me when I came upon it, but which turns out to be classic; when I mentioned the pattern to Ray Mayer he promptly directed my attention to a recent paper[6] in which the result is attributed to one Faulhaber (1615).

We have now in hand a population of *integer-valued functions of integers* which are (for evident reasons) known collectively as "figurate numbers." Such numbers acquire their quaint charm partly from the fact that they spring from ennumerative aspects of some pretty figures, and partly from the fact that they support a rich variety of pretty algebraic identities—relationships which are in most cases suggested by the figures themselves, and which prove useful for the same reasons, and to the same extent, that the associated figures are commonplace. This is not deep mathematics, but neither is it utterly devoid of interest; Euler returned to the subject repeatedly,[7] and it has recently been pressed into elegant service by John Conway[8] as a model-in-miniature of what mathematics is all about.

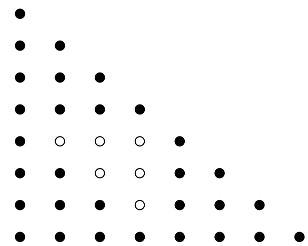
I digress to provide now a few examples of the "geometrically motivated algebraic identities" to which I just alluded. It is evident from the first of the following figures that

$$\Delta(2n) = 3\cdot\Delta(n) + \Delta(n-1)$$

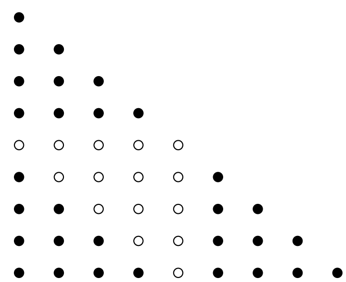
[6] A. F. Beardon, "Sums of powers of integers," Amer. Math.Monthly **103**, 201, (1996); Beardon's Theorem 3.1 appears to be precisely the result I had conjectured.

[7] See *Capitel 5: Von den Figurirten oder Vieleckigten Zahlen* §§425–439 of *Vollständige Anleitung zur Algebra, Erster Theil* (1770), which can be found at pp. 159–164 of *Leonhardi Euleri Opera Omnia, Series I, Volume I* (1911).

[8] See Chapter II of *The Book of Numbers* by Conway & R. K. Guy (1996).



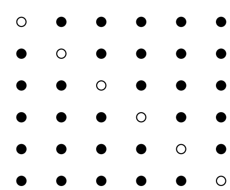
while from the following figure



we obtain an equation that describes not $\Delta(\text{even})$ but $\Delta(\text{odd})$:

$$\Delta(2n+1) = 3\cdot\Delta(n) + \Delta(n+1)$$

Again, it is evident—depending upon which way we read the following figure



—that

$$\begin{aligned} n^2 &= 2\cdot\Delta(n-1) + n \\ &= 2\cdot\Delta(n) - n \\ &= 2\cdot T(n,2) - T(n,1) \end{aligned}$$

of which $n^2 = 2\cdot\frac{1}{2}n(1+n) - n$ provides algebraic confirmation. Less obviously

$$\begin{aligned} n^3 &= 6\cdot T(n,3) - 6\cdot T(n,2) + T(n,1) \\ &= [n(1+n)(2+n)] - [3n(1+n)] + [n] \\ &= [n^3+3n^2+2n] - [3n^2+3n] + [n] \end{aligned}$$

$$\begin{aligned} n^4 &= 24\cdot T(n,4) - 36\cdot T(n,3) + 14\cdot T(n,2) - T(n,1) \\ &\;\;\vdots \end{aligned}$$

By adjustment of the preceding figure we obtain

$$\begin{array}{cccccc}
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\circ & \bullet & \bullet & \bullet & \bullet & \bullet \\
\circ & \circ & \bullet & \bullet & \bullet & \bullet \\
\circ & \circ & \circ & \bullet & \bullet & \bullet \\
\circ & \circ & \circ & \circ & \bullet & \bullet \\
\circ & \circ & \circ & \circ & \circ & \bullet
\end{array}$$

giving

$$n^2 = \Delta(n) + \Delta(n-1)$$

and by stacking such figures we obtain

$$S(n;2) = T(n;3) + T(n-1;3)$$

which in algebraic notation means

$$\begin{aligned}
\sum_{k=1}^{n} k^2 &= \tfrac{1}{6}n(1+n)(2+n) + \tfrac{1}{6}(n-1)n(1+n) \\
&= \tfrac{1}{6}n(1+n)(1+2n) \qquad (2)
\end{aligned}$$

and therefore reproduces a classic formula stated previously. Conway provides two alternative constructions of this result, and discusses also a great many variants of the same general theme.[9]

2. Constructing figurate number formulae by extrapolation from leading cases. I turn now to description of an idea which began as a doodle,[10] but will permit us to construct figurate number formulae (not by appeal to appropriately-drawn figures but) from leading-case data by means of a standardized algebraic procedure. That procedure will provide a kind of bridge to the methods of classical interpolation theory. To illustrate the procedure, we look back again to the triangular numbers; since

$$\Delta(n) \equiv \sum_{k=1}^{n} k \quad \text{becomes} \quad \int_0^n k\,dk = \tfrac{1}{2}n^2 \quad \text{in the continuous limit}$$

we infer that

$$\Delta(n) = a + bn + cn^2$$

We will use the data

$$\begin{aligned}
\Delta(0) &= 0 = a + b0 + c0^2 \\
\Delta(1) &= 1 = a + b1 + c1^2 \\
\Delta(2) &= 3 = a + b2 + c2^2
\end{aligned}$$

[9] Additional constuctions can be found in R. B. Nelson's *Proofs Without Words* (1993).

[10] It was late afternoon on a spring Tuesday in the early 1980's, and President Bragdon's Faculty Advisory Committee had become lost in pursuit of some long-forgotten fine point.

to fix the values of a, b and c. Immediately

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}$$

giving

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \tfrac{1}{2}\begin{pmatrix} 2 & 0 & 0 \\ -3 & 4 & -1 \\ 1 & -2 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix} = \tfrac{1}{2}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

and we recover the familiar result

$$\begin{aligned} \Delta(n) = \tfrac{1}{2}n + \tfrac{1}{2}n^2 &= \tfrac{1}{2}n(n+1) \\ &\sim \tfrac{1}{2}n^2 \quad \text{for } n \text{ large} \end{aligned}$$

Suppose our initial conjecture had on the other hand read

$$\Delta(n) = a + bn + cn^2 + dn^3$$

Proceeding exactly as before—but from an enlarged data set

$$\begin{aligned} \Delta(0) = 0 &= a + b0 + c0^2 + d0^3 \\ \Delta(1) = 1 &= a + b1 + c1^2 + d1^3 \\ \Delta(2) = 3 &= a + b2 + c2^2 + d2^3 \\ \Delta(3) = 6 &= a + b3 + c3^2 + d3^3 \end{aligned}$$

—we have

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 6 \end{pmatrix}$$

$$\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \tfrac{1}{6}\begin{pmatrix} 6 & 0 & 0 & 0 \\ -11 & 18 & -9 & 2 \\ 6 & -15 & 12 & -3 \\ -1 & 3 & -3 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ 3 \\ 6 \end{pmatrix} = \tfrac{1}{2}\begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

We have, in short, labored harder than necessary to reproduce precisely our previous result (or, from another point of view, to demonstrate that $\Delta(n)$ is in fact *quadratic* in n). Actually, we labored harder than necessary already in the first instance, for $\Delta(0) \Longrightarrow a = 0$, and obviates any need to include a in our list of undetermined coefficients. And it is only by convenient convention that we worked from "leading data;" we could equally well have worked from (say) $\Delta(5) = 15$ and $\Delta(14) = 105$.

Looking now to another example (the example which figured in my original doodle), we proceed from the observation that

$$S(n;2) \equiv \sum_{k=1}^{n} k^2 \quad \text{becomes} \quad \int_0^n k^2\,dk = \tfrac{1}{3}n^3 \quad \text{in the continuous limit}$$

to the conjecture that

$$S(n;2) = a + bn + cn^2 + dn^3$$

From

$$\begin{aligned} S(0;2) \quad &: \quad & 0 &= a + b0 + c0^2 + d0^3 \\ S(1;2) \quad &: \quad 1^2 = & 1 &= a + b1 + c1^2 + d1^3 \\ S(2;2) \quad &: \quad 1^2 + 2^2 = & 5 &= a + b2 + c2^2 + d2^3 \\ S(3;2) \quad &: \quad 1^2 + 2^2 + 3^2 = & 14 &= a + b3 + c3^2 + d3^3 \end{aligned}$$

we have

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 5 \\ 14 \end{pmatrix}$$

giving

$$\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \tfrac{1}{6} \begin{pmatrix} 6 & 0 & 0 & 0 \\ -11 & 18 & -9 & 2 \\ 6 & -15 & 12 & -3 \\ -1 & 3 & -3 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 5 \\ 14 \end{pmatrix} = \tfrac{1}{6} \begin{pmatrix} 0 \\ 1 \\ 3 \\ 2 \end{pmatrix}$$

whence again the classic formula

$$S(n;2) = \tfrac{1}{6}n + \tfrac{3}{6}n^2 + \tfrac{2}{6}n^3 = \tfrac{1}{6}n(1+n)(1+2n)$$

In each of the preceding examples the pattern of the argument has been the same: we require of a polynomial $P(x) = p_0 + p_1x + p_2x^2 + \cdots + p_nx^n$ that it assume designated values $\{P_0, P_1, P_2, \ldots, P_n\}$ at $x = \{0, 1, 2, \ldots, n\}$, and so write

$$\begin{pmatrix} 1 & 0 & 0^2 & \ldots & 0^n \\ 1 & 1 & 1^2 & \ldots & 1^n \\ 1 & 2 & 2^2 & \ldots & 2^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & n & n^2 & \ldots & n^n \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix} = \begin{pmatrix} P_0 \\ P_1 \\ P_2 \\ \vdots \\ P_n \end{pmatrix}$$

giving

$$\begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0^2 & \ldots & 0^n \\ 1 & 1 & 1^2 & \ldots & 1^n \\ 1 & 2 & 2^2 & \ldots & 2^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & n & n^2 & \ldots & n^n \end{pmatrix}^{-1} \begin{pmatrix} P_0 \\ P_1 \\ P_2 \\ \vdots \\ P_n \end{pmatrix}$$

Thus "by extrapolation" do low-order instances of any selected figurate number lead us to a formulæ descriptive of the general instances of such numbers. Thus also do we establish contact with the central idea of. . .

3. Classical interpolation theory. Given a function $f(x)$ we seek a polynomial

$$P(x) = p_0 + p_1x + p_2x^2 + \cdots + p_nx^n$$

which assumes values coincident with those of $f(x)$ at the $(n+1)$-tuple of points $\{x_0, x_1, x_2, \ldots, x_n\}$; we want it to be, in other words, the case that

$$P(x_i) \equiv p_0 + p_1 x_i + p_2 x_i^2 + \cdots + p_n x_i^n = f_i \equiv f(x_i)$$

and propose to use the simple function $P(x)$ to interpolate/extrapolate the f-data $\{f_i\}$. Immediately

$$\begin{pmatrix} 1 & x_0 & x_0^2 & \ldots & x_0^n \\ 1 & x_1 & x_1^2 & \ldots & x_1^n \\ 1 & x_2 & x_2^2 & \ldots & x_2^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^n \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix} = \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix} \tag{3}$$

gives

$$\begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix} = \begin{pmatrix} 1 & x_0 & x_0^2 & \ldots & x_0^n \\ 1 & x_1 & x_1^2 & \ldots & x_1^n \\ 1 & x_2 & x_2^2 & \ldots & x_2^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^n \end{pmatrix}^{-1} \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix}$$

which we agree to abbreviate

$$\mathbf{p} = \mathbb{V}^{-1}\mathbf{f}$$

My $\mathbb{V}$-notation derives from the circumstance that $\det \mathbb{V}$ is a classic object called the "Vandermonde determinant."[11] A surprisingly simple argument[12] gives

$$V \equiv \det \mathbb{V} = \prod_{i>j}(x_i - x_j)$$

Look, for example, to the case $n = 2$:

$$\begin{pmatrix} p_0 \\ p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{pmatrix}^{-1} \begin{pmatrix} f_0 \\ f_1 \\ f_2 \end{pmatrix}$$

[11] Cramer's Rule (which had been anticipated already by Maclauarin in 1748) is described in Chapter 2 of Gabriel Cramer's *Introduction à l'analyse des lignes corbes algébriques* (1750), and contains an implicit allusion to the concept of a determinant. But determinants—detached form any reference to systems of linear equations—were apparently first studied by A. Vandermone (1735-1796), whom Lebesgue has argued deserves to be called the "father of the theory of determinants." Vandermonde was known as a mathematician among musicians, but as a musician among mathematicians, and was active as a mathematician only between the years 1770 and 1774, when he produced a total of four papers. Those, however, anticipated aspects of the work of Abel and Galois, and also of Gauss, by whom they were well regarded. Vandermonde was in later years an active revolutionary and occasional musical theorist.

[12] See, for example, L. Weisner, *Introduction to the Theory of Equations* (1949), p. 56.

Mathematica confirms that $V = (x_2 - x_0)(x_2 - x_1)\cdot(x_1 - x_0)$ and supplies the information that

$$\begin{aligned}
p_0 &= +\frac{x_1 x_2}{(x_0-x_1)(x_0-x_2)}f_0 + \frac{x_0 x_2}{(x_1-x_0)(x_1-x_2)}f_1 + \frac{x_0 x_1}{(x_2-x_0)(x_2-x_1)}f_2\\
p_1 &= -\frac{x_1 + x_2}{(x_0-x_1)(x_0-x_2)}f_0 - \frac{x_0 + x_2}{(x_1-x_0)(x_1-x_2)}f_1 - \frac{x_0 + x_1}{(x_2-x_0)(x_2-x_1)}f_2\\
p_2 &= +\frac{1}{(x_0-x_1)(x_0-x_2)}f_0 + \frac{1}{(x_1-x_0)(x_1-x_2)}f_1 + \frac{1}{(x_2-x_0)(x_2-x_1)}f_2
\end{aligned}$$

So we have

$$f(x) \approx P(x)$$

with

$$\begin{aligned}
P(x) = &+\left\{\tfrac{x_1 x_2}{(x_0-x_1)(x_0-x_2)}f_0 + \tfrac{x_0 x_2}{(x_1-x_0)(x_1-x_2)}f_1 + \tfrac{x_0 x_1}{(x_2-x_0)(x_2-x_1)}f_2\right\}\\
&-\left\{\tfrac{x_1+x_2}{(x_0-x_1)(x_0-x_2)}f_0 + \tfrac{x_0+x_2}{(x_1-x_0)(x_1-x_2)}f_1 + \tfrac{x_0+x_1}{(x_2-x_0)(x_2-x_1)}f_2\right\}x\\
&+\left\{\tfrac{1}{(x_0-x_1)(x_0-x_2)}f_0 + \tfrac{1}{(x_1-x_0)(x_1-x_2)}f_1 + \tfrac{1}{(x_2-x_0)(x_2-x_1)}f_2\right\}x^2
\end{aligned}$$

which by mere reorganization becomes

$$P(x) = f_0\cdot L_0(x) + f_1\cdot L_1(x) + f_2\cdot L_2(x) \tag{4}$$

with

$$\begin{aligned}
L_0(x) &\equiv \frac{x_1 x_2}{(x_0-x_1)(x_0-x_2)} - \frac{x_1+x_2}{(x_0-x_1)(x_0-x_2)}x + \frac{1}{(x_0-x_1)(x_0-x_2)}x^2\\
L_1(x) &\equiv \frac{x_0 x_2}{(x_1-x_0)(x_1-x_2)} - \frac{x_0+x_2}{(x_1-x_0)(x_1-x_2)}x + \frac{1}{(x_1-x_0)(x_1-x_2)}x^2\\
L_2(x) &\equiv \frac{x_0 x_1}{(x_2-x_0)(x_2-x_1)} - \frac{x_0+x_1}{(x_2-x_0)(x_2-x_1)}x + \frac{1}{(x_2-x_0)(x_2-x_1)}x^2
\end{aligned}$$

Lagrange observed that (4) can be deduced by an argument that makes no explicit appeal to the methods of linear algebra, and can at the same time be expressed much more compactly/memorably. He begins by introducing

$$\begin{aligned}
L(x) &\equiv (x-x_0)(x-x_1)(x-x_2) \\
&= \text{polynomial with roots at the data points } \{x_0, x_1, x_2\}
\end{aligned}\tag{5}$$

Then

$$L'(x) = (x-x_1)(x-x_2) + (x-x_0)(x-x_2) + (x-x_0)(x-x_1)$$

gives

$$\begin{aligned}
L'(x_0) &= (x_0-x_1)(x_0-x_2)\\
L'(x_1) &= (x_1-x_0)(x_1-x_2)\\
L'(x_2) &= (x_2-x_0)(x_2-x_1)
\end{aligned}$$

from which we get the denominators. Moreover

$$\begin{aligned}
\frac{L(x)}{(x-x_0)} &= (x-x_1)(x-x_2) = x_1x_2 - (x_1+x_2)x + x^2\\
\frac{L(x)}{(x-x_1)} &= (x-x_0)(x-x_2) = x_0x_2 - (x_0+x_2)x + x^2\\
\frac{L(x)}{(x-x_2)} &= (x-x_0)(x-x_1) = x_0x_1 - (x_0+x_1)x + x^2
\end{aligned}$$

from which we get the numerators. The polynomial (3) can therefore be described

$$\begin{aligned}
P(x) &= f_0\cdot\frac{L(x)}{(x-x_0)L'(x_0)} + f_1\cdot\frac{L(x)}{(x-x_1)L'(x_1)} + f_2\cdot\frac{L(x)}{(x-x_2)L'(x_2)}\\
&= \sum_k f_k\cdot\frac{L(x)}{(x-x_k)L'(x_k)} \;:\; \textbf{Lagrange interpolation formula} \qquad (6)
\end{aligned}$$

The $P(x)$ thus constructed is a polynomial, but—owing to the structure of the denominators—is not manifestly so. According to K. Rektorys[13] "the Lagrange interpolation formula, although very important in theoretical considerations, is not…suitable for numerical evaluation." Rektorys turns therefore (as now also do I) to an account of **Newton's interpolation technique**. From the data

$$f_0 \equiv f(x_0) \quad f_1 \equiv f(x_1) \quad f_2 \equiv f(x_2)$$

Newton forms[14] ascending generations of "divided differences"

$$f(x_0,x_1) \equiv \frac{f_1-f_0}{x_1-x_0} \qquad f(x_2,x_1) \equiv \frac{f_2-f_1}{x_2-x_1}$$

$$f(x_0,x_1,x_2) \equiv \frac{f(x_2,x_1)-f(x_1,x_0)}{x_2-x_0}$$

and writes

$$\begin{aligned}
P(x) &= f_0 + (x-x_0)f(x_0,x_1) + (x-x_0)(x-x_1)f(x_0,x_1,x_2) \qquad (7)\\
&= f_0 + (x-x_0)\frac{f_1-f_0}{x_1-x_0}\\
&\qquad + (x-x_0)(x-x_1)\frac{(f_2-f_1)(x_1-x_0)-(f_1-f_0)(x_2-x_1)}{(x_2-x_0)(x_2-x_1)(x_1-x_0)}\\
&= f_0\left\{1 - \frac{x\;-x_0}{x_1-x_0} + \frac{(x\;-x_0)(x\;-x_1)(x_2-x_1)}{(x_2-x_0)(x_2-x_1)(x_1-x_0)}\right\} + \text{etc.}\\
&= f_0\left\{\frac{x_1x_2-(x_1+x_2)x+x^2}{(x_0-x_1)(x_0-x_2)}\right\} + \text{etc.}
\end{aligned}$$

[13] *Survey of Applicable Mathematics* (1969), p. 1223.
[14] See Abramowitz & Stegun, p. 877.

Evidently Newton's (7) and Lagrange's (6) comprise merely distinct modes of approaching and displaying the identically same result. A great number of other organizational principles are described in the literature; all amount, in effect, to expansion of a determinant, since (by a fundamental property which Vandermonde was the first to identify)

$$\det\begin{pmatrix} P(x) & 1 & x & x^2 & \dots & x^n \\ f_0 & 1 & x_0 & x_0^2 & \dots & x_0^n \\ f_1 & 1 & x_1 & x_1^2 & \dots & x_1^n \\ f_2 & 1 & x_2 & x_2^2 & \dots & x_2^n \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ f_n & 1 & x_n & x_n^2 & \dots & x_n^n \end{pmatrix} = 0$$

forces $P(x_k) = f_k : k = 0, 1, \dots, n$.

In a broad class of practical applications one has special interest in (i) optimizing the selection of the sample points x_k and (ii) estimating/minimizing the *error* inherent in the interpolation procedure. For when we write

$$f(x) \approx P(x)$$

we mean that

$$f(x) = P(x) + R(x) \text{ and } R(x) \text{ is in some absolute/relative sense “small”}$$

The question is: “How small is ‘small’?” To address the question one must, of course, know something about the distinguishing properties of the function $f(x)$, and it is in reference to those properties (whatever they are) that the answer will be phrased. For example, if $f(x)$ is known to be $n+1$ times differentiable on the interval $[a, b]$ where

$$\begin{aligned} a &\equiv \min\{x_0, x_1, x_2, \dots, x_n\} \\ b &\equiv \max\{x_0, x_1, x_2, \dots, x_n\} \end{aligned}$$

then it can be shown that

$$|R(x)| \le \tfrac{1}{(n+1)!}|(x-x_0)(x-x_1)\cdots(x-x_n)| \cdot \max|f^{n+1}(\xi)| \quad : \quad \xi \in [a, b]$$

My own present interest in interpolation theory lies, however, elsewhere, so I will not pursue this aspect of the subject.

4. Illustrative theoretical application of the Lagrange interpolation formula. If the real $N \times N$ matrix $\mathbb{A}$ is antisymmetric and if

$$\mathbb{R} \equiv e^{\mathbb{A}}$$

then $\mathbb{R}^{-1} = \mathbb{R}^{\mathsf{T}}$ and $\det \mathbb{R} = e^{\operatorname{tr}\mathbb{A}} = e^0 = +1$: $\mathbb{R}$ is a *(proper) rotation matrix*, and

$$\mathbf{x} \longrightarrow \mathbf{x}' = \mathbb{R}\mathbf{x}$$

serves to describe a (proper) rotation in N-dimensional space. A sharp sense of the geometrical meaning of such a transformation can be achieved by an argument which hinges on the spectral properties of $\mathbb{A}$ and which makes essential use of algebraic ideas we have come to associate with the Lagrange interpolation formula.[15]

By quick calculation

$$\begin{vmatrix} -\lambda & a \\ -a & -\lambda \end{vmatrix} = +\big\{a^2+\lambda^2\big\}$$
$$\begin{vmatrix} -\lambda & a & b \\ -a & -\lambda & c \\ -b & -c & -\lambda \end{vmatrix} = -\big\{(a^2+b^2+c^2)\lambda+\lambda^3\big\}$$

so

$$\mathbb{A}_{2\times2} \equiv \begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix} \quad \text{has eigenvalues } \{\pm i\varphi\} \text{ with } \varphi \equiv \sqrt{a^2}$$
$$\mathbb{A}_{3\times3} \equiv \begin{pmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{pmatrix} \quad \text{has eigenvalues } \{0, \pm i\varphi\} \text{ with } \varphi \equiv \sqrt{a^2+b^2+c^2}$$

Generally, the characteristic polynomial

$$\det(\mathbb{A}_\text{N} - \lambda\mathbb{I}\,) \text{ is } \begin{cases} \text{even when } N \text{ is even} \\ \text{odd when } N \text{ is odd} \end{cases}$$

but is in every case real, though its roots—the eigenvalues of $\mathbb{A}_\text{N}$—are in every case imaginary.[16] Evidently

$$\det(\mathbb{A}_\text{N} - \lambda\mathbb{I}\,) = (-)^N \begin{cases} (\lambda^2+\varphi_1^2)(\lambda^2+\varphi_2^2)\cdots(\lambda^2+\varphi_n^2) & \text{if } N=2n \\ \lambda(\lambda^2+\varphi_1^2)(\lambda^2+\varphi_2^2)\cdots(\lambda^2+\varphi_n^2) & \text{if } N=2n+1 \end{cases}$$

where the real numbers φ_k $(k=1,2,\ldots,n)$ serve to describe the (imaginary) spectrum of $\mathbb{A}$:

$$\text{eigenvalues of } \mathbb{A}_\text{N} = \begin{cases} \{\pm\lambda_1, \pm\lambda_2, \ldots, \pm\lambda_n\} & \text{if } N=2n \\ \{\lambda_0, \pm\lambda_1, \pm\lambda_2, \ldots, \pm\lambda_n\} & \text{if } N=2n+1 \end{cases}$$

with $\lambda_0 \equiv 0$ and $\lambda_k \equiv i\varphi_k$.

---

[15] Here I must be content merely to outline the argument; for details (and the proofs of some bald assertions), see "What does an N-dimensional rotation look like?" (Notes for a Reed College Math Seminar, presented 14 February 1980) in TRANSFORMATIONAL PHYSICS & PHYSICAL GEOMETRY (1971–1983).

[16] The real antisymmetry of $\mathbb{A}$ implies the hermiticity of $\mathbb{H} \equiv i\mathbb{A}$, and the eigenvalues of $\mathbb{H}$ are, by a familiar argument, necessarily real.

Looking now specifically to the $\boxed{\text{even-dimensional case } N = 2n}$: the Cayley-Hamilton theorem supplies

$$(\mathbb{A}^2 - \lambda_1^2\,\mathbb{I})(\mathbb{A}^2 - \lambda_2^2\,\mathbb{I})\cdots(\mathbb{A}^2 - \lambda_n^2\,\mathbb{I}) = \mathbb{O}$$

or again

$$(\mathbb{S} - s_1\,\mathbb{I})(\mathbb{S} - s_2\,\mathbb{I})\cdots(\mathbb{S} - s_n\,\mathbb{I}) = \mathbb{O}$$

where the notation has been designed to emphasize the real symmetry—whence the hermiticity—of the matrix $\mathbb{S} \equiv \mathbb{A}^2$, of which $\{s_1, s_2, \ldots, s_n\}$ are the (in the simplest case) doubly-degenerate eigenvalues. The spectral theory of hermitian operators now supplies

$$\mathbb{S} = s_1\mathbb{P}_1 + s_2\mathbb{P}_2 + \cdots + s_n\mathbb{P}_n \tag{8}$$

where $\{\mathbb{P}_1, \mathbb{P}_2, \ldots, \mathbb{P}_n\}$ is the complete orthogonal set of projection matrices

$$\begin{aligned} \mathbb{P}_k^2 = \mathbb{P}_k \quad &: \quad k = 1, 2, \ldots, n \\ \mathbb{P}_j\mathbb{P}_k = \mathbb{O} \quad &: \quad j \neq k \\ \mathbb{P}_1 + \mathbb{P}_2 + \cdots + \mathbb{P}_n &= \mathbb{I} \end{aligned}$$

which project onto (and serve thus to define) the orthogonal 2-dimensional eigenspaces of $\mathbb{S}$. Explicit descriptions of the matrices $\mathbb{P}_k$ will be developed in a moment, but from results already in hand it follows that

$$\left.\begin{array}{lllll} \mathbb{A}^0 & = & s_1^0\mathbb{P}_1 + & s_2^0\mathbb{P}_2 + \cdots + & s_n^0\mathbb{P}_n \\ \mathbb{A}^{2\nu} = \mathbb{S}^\nu & = & s_1^\nu\mathbb{P}_1 + & s_2^\nu\mathbb{P}_2 + \cdots + & s_n^\nu\mathbb{P}_n \\ & = & \lambda_1^{2\nu}\mathbb{P}_1 + & \lambda_2^{2\nu}\mathbb{P}_2 + \cdots + & \lambda_n^{2\nu}\mathbb{P}_n \\ \mathbb{A}^{2\nu+1} = \mathbb{A}\mathbb{S}^\nu & = & -i\lambda_1^{2\nu+1}\mathbb{A}_1\mathbb{P}_1 - & i\lambda_2^{2\nu+1}\mathbb{A}_2\mathbb{P}_2 - \cdots - & i\lambda_n^{2\nu+1}\mathbb{A}_n\mathbb{P}_n \end{array}\right\} \tag{9}$$

where $\nu = 1, 2, \ldots$ and the real matrices $\mathbb{A}_k$ are defined $\mathbb{A}_k \equiv i\mathbb{A}/\lambda_k = \mathbb{A}/\varphi_k$. From this information it follows, in particular, that

$$\begin{aligned} \mathbb{R} = e^{\mathbb{A}} &= \sum_{\nu=0}^{\infty}\left\{\frac{1}{(2\nu)!}\mathbb{A}^{2\nu} + \frac{1}{(2\nu+1)!}\mathbb{A}^{2\nu+1}\right\} \\ &= \sum_{k=1}^{n}\big(\cosh\lambda_k\cdot\mathbb{I} - i\sinh\lambda_k\cdot\mathbb{A}_k\big)\mathbb{P}_k \\ &= \sum_{k=1}^{n}\big(\cos\varphi_k\cdot\mathbb{I} + \sin\varphi_k\cdot\mathbb{A}_k\big)\mathbb{P}_k \end{aligned} \tag{10}$$

It is mainly to facilitate discussion of the simple geometrical interpretation of this result that we look now (as promised) to the explicit construction of the projection matrices $\mathbb{P}_k$; to write

$$\begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ s_1 & s_2 & s_3 & \dots & s_n \\ s_1^2 & s_2^2 & s_3^2 & \dots & s_n^2 \\ \vdots & \vdots & \vdots & & \vdots \\ s_1^{n-1} & s_2^{n-1} & s_3^{n-1} & \dots & s_n^{n-1} \end{pmatrix} \begin{pmatrix} \mathbb{P}_1 \\ \mathbb{P}_2 \\ \mathbb{P}_3 \\ \vdots \\ \mathbb{P}_n \end{pmatrix} = \begin{pmatrix} \mathbb{I} \\ \mathbb{S} \\ \mathbb{S}^2 \\ \vdots \\ \mathbb{S}^{n-1} \end{pmatrix} \tag{11}$$

is simply to renotate some leading case instances of an equation which at (9) was already displayed in the general case, but does serve usefully to put us in mind of algebraic material developed in §3, for (11) is structurally identical to (3) except in one detail: the "Vandermonde matrix" has been transposed. But from results already in hand—from

$$\begin{pmatrix} p_0 \\ p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{pmatrix}^{-1} \begin{pmatrix} f_0 \\ f_1 \\ f_2 \end{pmatrix}$$

$$\downarrow$$

$$= \text{result developed at the top of p. 12}$$

and the elementary fact that

$$(\text{transpose})^{-1} = (\text{inverse})^{\mathsf{T}}$$

—we have

$$\begin{pmatrix} p_0 \\ p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ x_0 & x_1 & x_2 \\ x_0^2 & x_1^2 & x_2^2 \end{pmatrix}^{-1} \begin{pmatrix} f_0 \\ f_1 \\ f_2 \end{pmatrix}$$

$$\downarrow$$

$$p_0 = \frac{x_1 x_2 f_0 - (x_1 + x_2) f_1 + f_2}{(x_0 - x_1)(x_0 - x_2)}$$

$$p_1 = \frac{x_0 x_2 f_0 - (x_0 + x_2) f_1 + f_2}{(x_1 - x_0)(x_1 - x_2)}$$

$$p_2 = \frac{x_0 x_1 f_0 - (x_0 + x_1) f_1 + f_2}{(x_2 - x_0)(x_2 - x_1)}$$

In cases of the type $f_k = x^k$ in which we at present have special interest the preceding equations simplify; they assume, in fact, precisely the structure of the previously-encountered Lagrange polynomials:

$$p_0 = L_0(x) = \frac{(x\ \ - x_1)(x\ \ - x_2)}{(x_0 - x_1)(x_0 - x_2)} = \frac{L(x)}{(x - x_0) L'(x_0)}\ \text{, etc.}$$

Thus is it seen to follow from (11) that

$$\mathbb{P}_k = L_k(\mathbb{S}) \quad \text{with} \quad L_k(\mathbb{S}) = \prod_{\substack{j=1 \\ j \neq k}}^{n} \frac{(\mathbb{S} - s_j \mathbb{I})}{(s_k - s_j)} = \prod_{\substack{j=1 \\ j \neq k}}^{n} \frac{(\,\mathbb{A}^2 + \varphi_j^2 \mathbb{I})}{(-\varphi_k^2 + \varphi_j^2)} \tag{12}$$

Drawing as needed upon the Cayley-Hamilton theorem, one either sees directly or can readily show that the matrices $\mathbb{P}_k$ are real symmetric matrices that commute with $\mathbb{A}$ and with each other, that each $\mathbb{P}_k$ projects onto a 2-space ($\mathbb{P}^2 = \mathbb{P}$ and $\det(\mathbb{P} - \lambda\mathbb{I}) \sim (\lambda - 1)^2\lambda^{n-2}$), and that collectively they comprise a complete commutative set of orthogonal projectors. Let the arbitrary N-vector $\mathbf{x}$ be decomposed

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_n \quad \text{with} \quad \mathbf{x}_k \equiv \mathbb{P}_k\,\mathbf{x} \in k^{\text{th}} \text{ eigenspace}$$

Let $\mathbf{a}_k \equiv \mathbb{A}\,\mathbf{x}_k$. From $\mathbb{P}_k\,\mathbf{a}_k = \mathbb{P}_k\mathbb{A}\,\mathbf{x}_k = \mathbb{A}\,\mathbb{P}_k\,\mathbf{x}_k = \mathbb{A}\,\mathbf{x}_k = \mathbf{a}_k$ we see that so also does $\mathbf{a}_k \in k^{\text{th}}$ eigenspace. From the antisymmetry of $\mathbb{A}$ it follows that $\mathbf{x}_k \perp \mathbf{a}_k$; i..e., that $\mathbf{x}_k \cdot \mathbf{a}_k = \mathbf{x}_k \cdot \mathbb{A}\,\mathbf{x}_k = 0$. Moreover,

$$\begin{aligned} \mathbf{a}_k \cdot \mathbf{a}_k = \mathbf{x}_k \cdot \mathbb{A}^{\mathsf{T}}\mathbb{A}\,\mathbf{x}_k &= -\mathbf{x}_k \cdot \mathbb{S}\mathbb{P}_k\,\mathbf{x}_k \\ &= -s_k\,\mathbf{x}_k \cdot \mathbf{x}_k \quad \text{by (12) and the Cayley-Hamilton theorem} \\ &= +\varphi_k^2\,\mathbf{x}_k \cdot \mathbf{x}_k \end{aligned}$$

so not only are $\mathbf{x}_k$ and $\mathbf{y}_k \equiv \mathbf{a}_k/\varphi_k = \mathbb{A}_k\mathbb{P}_k\mathbf{x}$ perpendicular; they have the same Euclidean length. Returning in the light of this result to (10), we see that the action of $\mathbb{R}$ can be described

$$\begin{aligned} \mathbf{x} \longrightarrow \mathbb{R}\,\mathbf{x} &= \sum_{k=1}^{n} (\cos\varphi_k \cdot \mathbf{x}_k + \sin\varphi_k \cdot \mathbf{y}_k) \\ &= \text{sum of } n \text{ orthogonal copies of } \mathrm{O}(2) \end{aligned}$$

To achieve completeness in the $\boxed{\text{odd-dimensional case } N = 2n+1}$ one must adjoin to the projection matrices $\{\mathbb{P}_1, \mathbb{P}_2, \ldots, \mathbb{P}_n\}$ a matrix

$$\mathbb{P}_0 = \mathbb{I} - (\mathbb{P}_1 + \mathbb{P}_2 + \cdots + \mathbb{P}_n)$$

which projects onto the "dangling 1-space" familiar in the case $N = 3$ as the invariant "axis" of the rotation. One obtains

$$\mathbb{R} = \mathbb{P}_0 + \sum_{k=1}^{n} \big(\cos\varphi_k \cdot \mathbb{I} + \sin\varphi_k \cdot \mathbb{A}_k\big)\mathbb{P}_k \tag{13}$$

where the "Lagrange construction" (12) serves as before to define the projection matrices $\mathbb{P}_k$.

It is interesting to observe that the algebraic methods of interpolation theory—methods associated particularly with the name of Lagrange—have been central to the preceding discussion even though it was addessed to a geometrical question devoid of any recognizably "interpolative" features.[17]

[17] For further evidence of the natural intrusion of interpolative formulæ into the "function theory of matrices" see Chapter V of P. Lancaster, *Theory of Matrices* (1969) and the allusion to the "Sylvester interpolation formula" which appears on p. 102 of R. Bellman's *Introduction to Matrix Analysis* (1970)

5. Curve-fitting & generalized matrix inversion. To begin at the simple beginning: two points determine a line, but three points determine a triple of lines, and N points determine a $\frac{1}{2}N(N-1)$-fold population of lines. The notion of a "line drawn through N points" is, in the general case (i.e., unless those points conform to a "collinearity condition"), therefore absurd on its face; on the other hand, we can readily imagine situations in which we are motivated to speak of the line which "best approximates" such an impossibility. To phrase the issue analytically, we write

$$y = a + bx \tag{14}$$

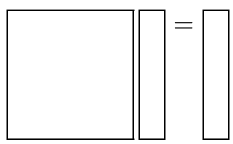
to describe lines in general, and to discover the line which interpolates between (x_1, y_1) and (x_2, y_2) we write

$$\begin{aligned} a + bx_1 &= y_1 \\ a + bx_2 &= y_2 \end{aligned}$$

which comprise a pair of conditions on a pair of undetermined parameters $\{a, b\}$. Since those parameters enter *linearly* into the construction of (14), we find it natural to write

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

and confront therefore a problem of type



(15)

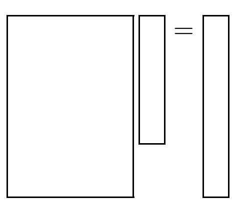
which is susceptible to analysis by the elementary methods of linear algebra. The N-point system of equations

$$\begin{aligned} a + bx_1 &= y_1 \\ a + bx_2 &= y_2 \\ &\vdots \\ a + bx_N &= y_N \end{aligned}$$

—which (appealing again to linearity) can be notated

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

and displays the structure



(16)

—is, on the other hand, insoluable because *over-determined*.

There are, in such a situation, several distinct ways in which we might proceed. For example, we might (in a phrase borrowed from the language of statistics) "modify our hypothesis:"

$$\begin{aligned} &y = a + bx \\ &\quad\downarrow \\ &y = \underbrace{a + bx + cx^2 + \cdots + qx^n}_{\text{number } n+1 \text{ of adjustable parameters } = \text{ number } N \text{ of data points}} \end{aligned} \tag{17}$$

To do so is to be led back again to the well-posed problem (15); we recover, in fact, a notational variant of (3), to which the classical methods of §3 directly pertain. We observe in this connection that (17)—which might be notated

$$\begin{aligned} y &= F(x; a, b, c, \ldots, q) \\ &= \text{linear combination of simple "power functions" } x^k \end{aligned} \tag{18}$$

—serves generally to inscribe on the $\{x, y\}$-plane not a line but a *curve*; the curve-fitting problem (15) acquires its linearity from the circumstance that the adjustable parameters $\{a, b, c, \ldots, q\}$ enter linearly into the construction of the function $F(x; a, b, c, \ldots, q)$, and that feature of the problem would remain intact if in place of (18) we were to write

$$= \text{linear combination of any prescribed functions } f_k(x) \tag{19}$$

In place of (3) we would then have

$$\begin{pmatrix} f_{11} & f_{12} & f_{13} & \cdots & f_{1n} \\ f_{21} & f_{22} & f_{23} & \cdots & f_{2n} \\ f_{31} & f_{32} & f_{33} & \cdots & f_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ f_{n1} & f_{n2} & f_{n3} & \cdots & f_{nn} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}$$

where $f_{ij} \equiv f_j(x_i)$ and where the adjustable parameters have now been notated $\{p_1, p_2, p_3, \ldots, p_n\}$. The generalization (18) $\mapsto$ (19) has entailed sacrifice of the polynomial-based material developed in §3, but has otherwise cost nothing; an instance of (15) has become another instance of (15).

In many contexts it would, however, be unacceptably alien to the spirit of the inquiry to "modify our hypothesis" every time we acquired an additional data point; our objective—at least in connection with the statistical analysis and interpretation of observational data—is to use a "least-possible number of adjustable parameters" to provide a "best-possible account of the available data," and we expect the acquisition of new data to result not in parameter proliferation but in reduced uncertainty. Looking in this light back again to the over-determined system (16), we observe that while it is (in the general case) not self-consistently possible to write

$$\begin{pmatrix} f_{11} & f_{12} & f_{13} & \dots & f_{1n} \\ f_{21} & f_{22} & f_{23} & \dots & f_{2n} \\ f_{31} & f_{32} & f_{33} & \dots & f_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ f_{m1} & f_{m2} & f_{m3} & \dots & f_{mn} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ \vdots \\ \vdots \\ y_m \end{pmatrix}$$

it is possible to set

$$\begin{pmatrix} f_{11} & f_{12} & f_{13} & \dots & f_{1n} \\ f_{21} & f_{22} & f_{23} & \dots & f_{2n} \\ f_{31} & f_{32} & f_{33} & \dots & f_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ f_{m1} & f_{m2} & f_{m3} & \dots & f_{mn} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_n \end{pmatrix} \equiv \begin{pmatrix} Y_1(p_1, p_2, \dots, p_n) \\ Y_2(p_1, p_2, \dots, p_n) \\ Y_3(p_1, p_2, \dots, p_n) \\ \vdots \\ \vdots \\ \vdots \\ Y_m(p_1, p_2, \dots, p_n) \end{pmatrix} \tag{20.1}$$

and to require of the parameters $\{p_1, p_2, \dots, p_n\}$ that

$$\sum_{i=1}^{m} \big[Y_i(p_1, p_2, \dots, p_n) - y_i\big]^2 = \text{minimum} \tag{20.2}$$

The sum of squares on the left can be notated

$$\begin{aligned} Q(\mathbf{p}) &\equiv (\mathbb{F}\mathbf{p} - \mathbf{y})^{\mathsf{T}}(\mathbb{F}\mathbf{p} - \mathbf{y}) \\ &= \mathbf{p}^{\mathsf{T}}\mathbb{S}\mathbf{p} - 2\mathbf{p}^{\mathsf{T}}\mathbf{q} + \mathbf{y}^{\mathsf{T}}\mathbf{y} \quad : \quad \mathbb{S} \equiv \mathbb{F}^{\mathsf{T}}\mathbb{F} \text{ is } n \times n \text{ symmetric and } \mathbf{q} \equiv \mathbb{F}^{\mathsf{T}}\mathbf{y} \\ &= \text{quadratic form in the variables } \mathbf{p}\text{, minimized at } \mathbf{p} = \mathbf{p}_{\text{minimal}} \end{aligned}$$

Immediately $\boldsymbol{\nabla}_{\!p} Q = 2(\mathbb{S}\,\mathbf{p} - \mathbf{q})$, so $\boldsymbol{\nabla}_{\!p} Q = \mathbf{0}$ entails

$$\begin{aligned} \mathbf{p}_{\text{minimal}} &= \mathbb{S}^{-1}\mathbf{q} = (\mathbb{F}^{\mathsf{T}}\mathbb{F})^{-1}\mathbb{F}^{\mathsf{T}}\mathbf{y} \\ &= \mathbb{E}\mathbf{y} \quad \text{where } \mathbb{E} \equiv (\mathbb{F}^{\mathsf{T}}\mathbb{F})^{-1}\mathbb{F}^{\mathsf{T}} \\ &\qquad\qquad \equiv n \times m \text{ “generalized left inverse” of } \mathbb{F} \end{aligned} \tag{21}$$

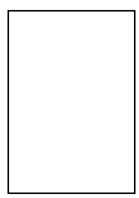
which, by the way, gives

$$Q(\mathbf{p}_{\text{minimal}}) = \mathbf{y}^{\mathsf{T}}(\mathbb{I} - \mathbb{F}\mathbb{E})\mathbf{y} \tag{22}$$

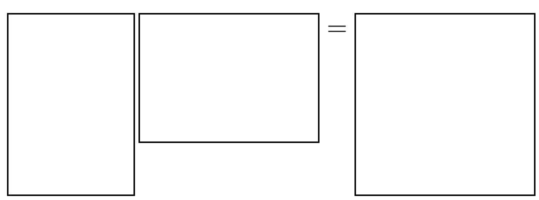
Several comments are now in order:

At (20) we have ventured down a path first explored in print (under the title "Nouvelles méthodes pour la détermination des orbites des comètes") by Legendre in 1805, though Gauss reported in his *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solum Ambientium* ("Theory of the Motion of Heavenly Bodies Moving about the Sun in Conic Sections," 1809) that he had been making routine use of the "Method of Least Squares" since 1795. Both Legendre and Gauss were concerned (or so we infer from the titles of their publications) with instances of the curve-fitting problem; i.e., with astrophysical applications of (20), but by the last decade of the 19$^{\text{th}}$ Century the method of least squares—which takes its name, obviously, from the structure of the "goodness-of-fit criterion" (20.2)—had become fundamental to a field which has relatively little to do with curve-fitting *per se*: it had become fundamental to the new discipline of *statistics*, where it had spawned such concepts as "linear regression" and "correlation coefficient." What began as a method for achieving the "imperfect best approximation" to the solution of a class of problems involving what might be called "analytic geometry in the presence of errors" had evolved—not too surprisingly, I think—into a "geometrical theory of errors."[18]

The method of least squares was originally presented as an exercise in applied calculus, not as an exercise in applied linear algebra (which in 1805 hadn't been invented yet!). When formulated in the latter terms it leads to a seldom-remarked *generalization of the concept of matrix inversion*, and it is to that implication of (20) that I now turn. Let $\mathbb{F}$ be an arbitrary $m\times n$ matrix, with $m\geq n$:

$\mathbb{F}\equiv$ 

The associated right-inversion problem—the problem of exhibiting an $n\times m$ matrix $\mathbb{G}$ such that

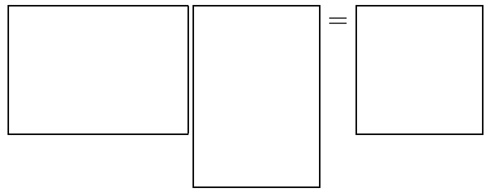
$\mathbb{F}\mathbb{G}=\mathbb{I}_{m\times m}\quad:$ 

—is (unless $m=n$) over-determined, since it imposes m^2 conditions on the $mn<m^2$ elements of $\mathbb{G}$. On the other hand, the left inversion problem—the

[18] For a nice account of this interesting chapter in the history of mathematics, see S. M. Stigler, *The History of Statistics; The Measurement of Uncertainty before 1900* (1986). Of course, (20) does continue to serve the curve-fitting purpose for which it was originally intended; see §§**1.6.6** & **3.8** and the entry `Fit`[*data, funs, vars*] on p. 1088 of *The Mathematica Book* (Version 3.0, 1996).

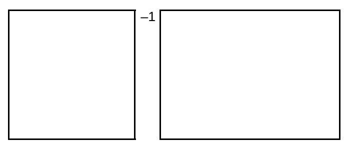
problem of exhibiting an $n \times m$ matrix $\mathbb{E}$ such that

$$\mathbb{E}\mathbb{F} = \mathbb{I}_{n\times n} \quad :$$



—is (unless $m = n$) under-determined, since it imposes only n^2 conditions on the $mn > n^2$ elements of $\mathbb{E}$; it admits, therefore, of many solutions. At (21) the method of least squares assigned special significance to one of those:

$$\mathbb{E} \equiv (\mathbb{F}^{\mathsf{T}}\mathbb{F})^{-1}\mathbb{F}^{\mathsf{T}} = \tag{23}$$



The presumption here is that the $n \times n$ square matrix $\mathbb{F}^{\mathsf{T}}\mathbb{F}$ is non-singular. Gratifyingly, (23) gives back the ordinary inverse $\mathbb{F}^{-1}$ when $\mathbb{F}$ itself is square ($m = n$) and non-singular. When that is the case—i.e., when the left inverse $\mathbb{E}$ is also a right inverse—the expression on the right side of (22) vanishes.

We have touched here on an idea which appears to have occurred first to E. H. Moore, whose remarks at a regional meeting in 1920 of the American Mathematical Society are summarized in that society's Bulletin[19] but attracted little attention. The subject was independently reinvented in the mid-1950s by Roger Penrose, whose initial publication[20] lacked clear motivation and was phrased quite abstractly, but was followed promptly by a paper[21] intended to establish "relevance to the statistical problem of finding 'best' approximate solutions of inconsistent systems of equations by the method of least squares." What I have called the "generalized left inverse" is sometimes called the "Moore-Penrose inverse," but is known to *Mathematica* as "`PsuedoInverse`[m]," and is considered by *Mathematica* to arise by specialized application of the so-called "singular value decomposition" of $\mathbb{F}$. For a good modern account of the subject and its applications, see S. L. Campbell & C. D. Meyer, *Generalized Inverses of Linear Transformations* (1979). Also useful is Appendix 2 in P. Lancaster's *Theory of Matrices* (1969) and the material on p.105 in Richard Bellman's *Introduction to Matrix Analysis* (1970).

---

[19] Bull. Amer. Math. Soc. (2) **26**, 394 (1920).

[20] "A generalized inverse for matrices," Proc. Camb. Phil. Soc. **51**, 406 (1955).

[21] "On best approximate solutions of linear matrix equations," Proc. Camb. Phil. Soc. **52**, 17 (1956).

6. Sums of powers. In §1 we introduced the functions

$$S(n;p) \equiv \sum_{k=1}^{n} k^p$$

that describe sums of powers, and found in particular that

$$\begin{aligned}
S(n;0) &= n \\
S(n;1) &= \tfrac{1}{2}n(1+n) \\
S(n;2) &= \tfrac{1}{6}n(1+n)(1+2n) \\
S(n;3) &= \tfrac{1}{4}n^2(1+n)^2 \\
&\vdots \\
S(n;p) &= \text{polynomial of order } p+1 \text{ in } n
\end{aligned} \tag{24}$$

These formulæ assign natural values to expressions of (for example) the type $S(n+\frac{1}{2};p)$, but do not clarify what we might mean by $S(n;p+\frac{1}{2})$. In an (ill-fated) effort to get a handle on that problem, we construct the generating function

$$\begin{aligned}
G(t;n) \equiv \sum_{p=0}^{\infty} \frac{1}{p!} S(n;p)t^p &= \sum_{p=0}^{\infty} \frac{1}{p!} \Big\{ \sum_{k=1}^{n} k^p \Big\} t^p \\
&= \sum_{k=1}^{n} e^{kt} \\
&= \sum_{k=1}^{n} (e^t)^k \\
&= \frac{e^{(n+1)t} - e^t}{e^t - 1}
\end{aligned}$$

Recalling now[22] the definition

$$\frac{te^{xt}}{e^t - 1} = \sum_{n=0}^{\infty} \frac{1}{n!} B_n(x)\, t^n$$

of the Bernoulli polynomials $B_n(x)$, we have

$$\begin{aligned}
G(t;n) &= \sum_{p=0}^{\infty} \frac{1}{p!} \Big\{ B_p(n+1) - B_p(1) \Big\} t^{p-1} \\
&= \sum_{p=0}^{\infty} \frac{1}{p!} \Big\{ \frac{B_{p+1}(n+1) - B_{p+1}(1)}{p+1} \Big\} t^p \quad \text{by } B_0(x) = 1
\end{aligned}$$

[22] See J. Spanier & K. Oldham, *Atlas of Functions* (1987), Chapter 19.

Finally we use $B_n(1) = (-)^n B_n$ (here B_n is the n^{th} Bernoulli number: see p. 38 of Spanier & Oldham) to obtain the famous identity

$$S(n;p) \equiv \sum_{k=1}^{n} k^p = \tfrac{1}{p+1}\Big\{B_{p+1}(n+1) + (-)^p B_{p+1}\Big\} \tag{25}$$

This formula makes explicit the meaning of (24), but does still not assign meaning to $1^p + 2^p + \cdots + n^p$ when p is non-integral. However, we have

$$S(n;p) = \Big(\frac{d}{dt}\Big)^p G(t;n)\Big|_{t=0} = \Big(\frac{d}{dt}\Big)^p \Big\{\frac{e^{(n+1)t} - e^t}{e^t - 1}\Big\}\Big|_{t=0} \quad : \quad p = 0, 1, 2, \ldots$$

which, in fact, works; *Mathematica* supplies

$$\begin{aligned}
\texttt{Limit[}\tfrac{e^{(n+1)t}-e^t}{e^t-1}\texttt{, t->0]} &= n\\
\texttt{Limit[}\partial_t\tfrac{e^{(n+1)t}-e^t}{e^t-1}\texttt{, t->0]} &= \tfrac{1}{2}n(1+n)\\
\texttt{Limit[}\partial_t\partial_t\tfrac{e^{(n+1)t}-e^t}{e^t-1}\texttt{, t->0]} &= \tfrac{1}{6}n(1+3n+2n^2)\\
\texttt{Limit[}\partial_t\partial_t\partial_t\tfrac{e^{(n+1)t}-e^t}{e^t-1}\texttt{, t->0]} &= \tfrac{1}{4}n^2(1+n)^2\\
&\vdots
\end{aligned}$$

from which we do in fact recover (24). The question now arises: Does it make any sense to write (say)

$$S(n,\tfrac{1}{2}) = 1^{\frac{1}{2}} + 2^{\frac{1}{2}} + \cdots + n^{\frac{1}{2}} = \Big(\frac{d}{dt}\Big)^{\frac{1}{2}} \Big\{\frac{e^{(n+1)t} - e^t}{e^t - 1}\Big\}\Big|_{t=0} \quad ?$$

The answer, disappointingly but clearly, is "No; generating functions—for the simplest of reasons—do not work that way." For if

$$G(t) = G_0 + G_1 t + \tfrac{1}{2!}G_2 t^2 + \cdots + \tfrac{1}{n!}G_n t^n + \cdots$$

then (as was familiar already to Maclaurin)

$$\big(\tfrac{d}{dt}\big)^p G(t)\big|_{t=0} = G_p \quad : \quad p = 0, 1, 2, \ldots$$

but

$$\big(\tfrac{d}{dt}\big)^{\frac{1}{2}} G(t) = \underbrace{G_0 \tfrac{1}{\sqrt{\pi t}}}_{\infty\text{ in the limit } t\downarrow 0} + G_1 2\sqrt{\tfrac{t}{\pi}} + G_2 \tfrac{4}{3}\sqrt{\tfrac{t^3}{\pi}} + \cdots + G_n \tfrac{1}{\Gamma(n+\frac{1}{2})} t^{n-\frac{1}{2}} + \cdots$$

Proceeding now formally—with no other objective than to avoid infinity on the one hand and triviality on the other—we might form

$$H(t) \equiv \sqrt{t}\cdot G(t) = G_0 t^{\frac{1}{2}} + G_1 t^{\frac{3}{2}} + \tfrac{1}{2!}G_2 t^{\frac{5}{2}} + \cdots + \tfrac{1}{n!}G_n t^{n+\frac{1}{2}} + \cdots$$

giving

$$\left(\tfrac{d}{dt}\right)^{\frac{1}{2}} H(t) = \underbrace{G_0\tfrac{1}{2}\sqrt{\pi}}_{=\,G_{\frac{1}{2}}} + G_1\tfrac{4}{3}\sqrt{\pi}\,t + \tfrac{1}{2!}G_2\tfrac{15}{16}\sqrt{\pi}\,t^2 + \cdots + \tfrac{1}{n!}G_n\tfrac{\Gamma(n+\frac{3}{2})}{\Gamma(n+1)}t^n + \cdots$$

Such a claim is, however, inconsistent with the facts of the matter in the particular case at hand[23] and is anyway implausible on its face: $G_{\frac{1}{2}}$ is trying to *interpolate* between G_0 and G_1, and should therefore be sensitive to the values of *all* the numbers $\{G_0, G_1, G_2, \ldots\}$, but to write $G_{\frac{1}{2}} = \frac{\sqrt{\pi}}{2}G_0$ is to assert that $G_{\frac{1}{2}}$ stands in a fixed/universal relationship to G_0, and is *independent* of the values assumed by $\{G_1, G_2, \ldots\}$. We conclude that *the generating function technique is ill-adapted to the problem of "interpolation in the exponent."*

[23] It is manifestly *not* the case that $S(n;0) = \frac{\sqrt{\pi}}{2}S(n;\frac{1}{2})$.