The third edition (1820), with additions and supplements, appears as Vol. 7, *Oeuvres complètes des Laplace*. First edition reprinted 1967, Brussels: Culture et Civilisation. A relevant portion is translated in Smith 1929, 588-604.

Legendre, A.M. 1805 *Nouvelles méthodes pour la détermination des orbits des comètes*. Paris. Appendix "Sur la méthode des moindres quarrés," translated in Smith 1929, 576-579.

Merriman, M. 1877 "A list of writings relating to the method of least squares, with historical and critical notes," *Transactions of the Connecticut Academy of Arts and Sciences* 4, 151-232.

Plackett, R. 1972 "Studies in the history of probability and statistics XXIX. The discovery of the method of least squares," *Biometrika* 59, 239-251.

Rabinovitch, N.L. 1973 *Probability and Statistical Inference in Ancient and Medieval Jewish Literature*. University of Toronto Press.

Sheynin, O.B. 1973 "Mathematical treatment of astronomical observations, a historical essay," *Arch Hist Ex Sci* 11, 97-126.

Smith, D.E. 1929 *A Source Book in Mathematics*. McGraw-Hill, New York. Reprinted in 2 volumes by Dover Publications, New York, 1959.

Struik, D.J. 1972 "Gergonne, Joseph Diaz [sic]," *Dictionary of Scientific Biography* (ed. C.C. Gillispie), Vol. 5, 367-369. Charles Scribner's Sons, New York.

Todhunter, I. 1865 *A History of the Mathematical Theory of Probability*. Reprinted by Chelsea Pub., New York, 1949 and 1965.

THE APPLICATION OF THE METHOD OF LEAST SQUARES TO THE INTERPOLATION OF SEQUENCES

By J.D. Gergonne

Translated by Ralph St. John, Bowling Green State University and S.M. Stigler, University of Wisconsin

Translators' Note: *An effort has been made not to introduce any modern* statistical *terminology and to reflect Gergonne's thinking accurately. To ease the way for modern readers, however, some of the* mathematical *terminology has been updated (examples: "polynomial function" for "fonction complète, rationelle et entière" and "derivatives" for "coefficiens différentiels." All italics, including those in the quotation from Laplace, are Gergonne's, as are the footnotes unless otherwise indicated. Some readers may be unfamiliar with the osculating circle, a geometric measure* *of curvature at a point and geometric analogue of a second derivative.*

When a function of a single variable is known, we can always determine rigorously and directly the values of the function and of its various derivatives at a given value of the independent variable. Similarly, given a curve we can always, for any abscissa, obtain the ordinate, the tangent, the osculating circle, etc.

Just as instead of giving a curve we can give only a certain number of its points, we can similarly instead of giving a function of a variable give only the values this function takes for a certain number of values of the independent variable, and subsequently ask what are the values of the function and its various derivatives for any other value of this variable. Similarly we could ask for a given abscissa what are the ordinate, the tangent, the osculating circle, etc. of a curve about which we know only a certain number of points. This constitutes the problem of the *interpolation of sequences.*

This problem obviously reduces to recovering from the given values, the function from which they were obtained, or from the given points, the plot of the curve on which we assume they are located. However, the problem is indeterminate for, given non-consecutive points, even an infinite number of them, we can always pass through them an infinity of different curves.*

These curves could very well differ notably from one another in certain parts of their range; the same difference will be observed in the ordinates, tangents, osculating circles, etc. for a given abscissa. However, we note that if the given points are close enough to each other, the curves which include them will not differ greatly over this interval, at least if none of the curves has within this interval an asymptote parallel to the axis of the ordinates. We also note that these given points can always be numerous enough, and, at the same time, sufficiently close to each other, that the differences between these curves become almost indistinguishable. The ordinates which result from a single abscissa within this range will therefore be essentially equal; however, the difference between the tangents can be more sensitive, that between the osculating circles even more so, and so forth.

We conclude from this that, if functions of diverse form have

---

* We can consult on this subject a dissertation on page 252 of volume 5 of this journal. [Trans. note: The article referred to, "Considerations philosophiques sur l'interpolation," is by Gergonne but contains no material relevant to statistics.]

the same value for certain known neighboring values of the independent variable without becoming infinite for any value included between these, then these functions will take on values scarcely different for other values of the independent variable included within the above interval. However, this will not be the case for the derivatives of the various functions, which can differ more and more as the corresponding order increases.

We can therefore, without noticeable error, arbitrarily adopt one of these functions as the desired function; similarly when many curves which pass through the same points have only slight differences, we can assume that any one of these is really the curve on which these points lie.

Since the curve or the function can be selected in an infinity of different ways, it is convenient to select the simplest way, that is, the parabolic curve or the polynomial function that graphically represents it. This choice is well founded since it is known that all finite functions of a finite variable can always be expressed in a series of increasing powers of this variable.

The procedure we have just arrived at is also that which is commonly followed; we assume that the ordinate of the desired curve is a polynomial function of the abscissa, into which we allow as many terms as there are sets of given values; the coefficients of these terms are unknown, and we determine them by assuming that the curve passes through the given points. Once these coefficients are determined, it is a simple matter to calculate the ordinate and the derivatives for any abscissa. However, we can rely on the values obtained from this formula only when it is applied to an abscissa within the interval containing the given points, and also not too close to the largest or the smallest.

This method, which was employed by Mr. Laplace in his memoir *Recherche des orbites des comètes*,* contains a source of error in the supposition that the curve is a parabolic curve. Nevertheless, if we could rigorously believe in the given values of the function, and if these values were very numerous and very close to each other, then what we have said above shows that the error resulting from this supposition would never be very large.

However, this is not always the case. The discrete values of the function, which we have used to construct our formula, are often deduced from experience or from observations subject to limited precision. Thus, as Mr. Legendre has observed,** it often happens that the errors which affect these observations can have more and more influence on the final solution and on the results we deduce from this solution, as more and more values are obtained.

---

* See the *Mémoires de l'Académie des Sciences*, Paris, for 1780.

** See his *Nouvelles méthodes pour la détermination des orbites des comètes*, Paris, 1806, p. iv.

Assume that we have plotted an arbitrary curve, and that we have obtained from it many ordinates very close to each other. Suppose we have subjected these ordinates to very small changes, sometimes positive and sometimes negative, and subsequently we attempt to pass a continuous curve through these altered ordinates. We will easily see that, even if these alterations have had only a very small influence on the size of intermediate ordinates, that is not the case with regard to the tangent, which may have undergone a notable change for the same abscissa, and this change may be even more noticeable with regard to the osculating circle.

These graphical observations can easily be confirmed by calculations. Suppose we have an odd number of given ordinates corresponding to equidistant abscissas, and assume that this common distance is one. Let zero be the abscissa and b the ordinate at the middle; 1, 2, 3 ... the abscissas and b^1, b^2, b^3... the ordinates which follow; -1, -2, -3, the abscissas and b_1, b_2, b_3,... the ordinates which precede. We wish to obtain the various derivates at zero. We obtain for the case of three ordinates

$$\frac{dy}{dx} = \frac{b^1-b_1}{2},\ \frac{d^2y}{dx^2} = (b^1+b_1)-2b;$$

for the case of five ordinates

$$\frac{dy}{dx} = \frac{8(b^1-b_1)-(b^2-b_2)}{12},\ \frac{d^2y}{dx^2} = -\frac{30b-16(b^1+b_2)+(b^2+b_2)}{12};$$

for the case of seven ordinates

$$\frac{dy}{dx} = \frac{45(b^1-b_1)-9(b^2-b_2)+(b^3-b_3)}{60},$$

$$\frac{d^2y}{dx^2} = -\frac{490b-270(b^1+b_1)+27(b^2+b_2)-2(b^3+b_3)}{180};$$

and so forth.

Suppose that the other ordinates are exact and that the ordinate b^1 is in error by the quantity β. Let $E\frac{dy}{dx}$, $E\frac{d^2y}{dx^2}$ denote the resulting errors in the derivatives at zero. It is easy to see that, in the case of three ordinates

$$E\frac{dy}{dx} = \frac{1}{2}\beta,\ E\frac{d^2y}{dx^2} = \frac{2}{2}\beta;$$

in the case of five ordinates

$$E\,\frac{dy}{dx} = \frac{2}{3}\beta,\quad E\,\frac{d^2y}{dx^2} = \frac{4}{3}\beta;$$

in the case of seven ordinates

$$E\,\frac{dy}{dx} = \frac{3}{4}\beta,\quad E\,\frac{d^2y}{dx^2} = \frac{6}{4}\beta.$$

Therefore the errors in the first order derivative increase as do the numbers $\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5},\ldots$ and thus tend monotonically to the actual error in the ordinate b^1. Similarly, the error in the second order derivative is double that of the first order derivative.

Mr. Legendre was therefore justified in saying that in increasing the number of values, we exposed ourselves to an increase in the errors in the same proportion. This result assumes that there is only one incorrect ordinate, which excludes all possibility of compensating errors. Moreover, this assumes that the incorrect ordinate is precisely that whose value, exact or not, exerts the most influence on our two derivatives.

Whatever the case, this source of error did not escape the attention of Mr. Laplace. Here are his comments [*Mécanique céleste*, Tome I, p. 201]*: "These expressions are more precise as there are more observations, and as the interval separating them is smaller. We could therefore use all the neighboring observations for the chosen period, if they were exact, but the errors to which they are subject would lead us to a false result. Therefore, *to reduce the influence of these errors, we must increase the interval of the extreme observations as we employ more observations*."

It would probably be more correct to say that *we must employ observations more and more distant from each other as we employ more observations*. We shall see, in effect, that with this procedure we can reduce these errors. Let a be the interval, assumed constant, which separates consecutive values of x, an interval which we previously assumed to be one. Our previous results then become, for three observations

$$E\,\frac{dy}{dx} = \frac{1}{2}\,\frac{\beta}{a},\quad E\,\frac{d^2y}{dx^2} = \frac{2}{2}\,\frac{\beta}{a^2}\;;$$

for five observations

---

* Trans note: This passage may be found on p. 411 of volume I of Bowditch's translation.

$$E\,\frac{dy}{dx} = \frac{2}{3}\,\frac{\beta}{a},\quad E\,\frac{d^2y}{dx^2} = \frac{4}{3}\,\frac{\beta}{a^2};$$

for seven observations

$$E\,\frac{dy}{dx} = \frac{3}{4}\,\frac{\beta}{a},\quad E\,\frac{d^2y}{dx^2} = \frac{6}{4}\,\frac{\beta}{a^2}$$

Therefore, as long as a takes on values that increase more rapidly than does the sequence $\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots$ our errors will continually decrease as we have more and more observations. Suppose, for example, that we increase the value of a according to the positive intergers. Let this value be one for the case of three observations. We thus have for three observations

$$E\,\frac{dy}{dx} = \frac{1}{2}\beta,\quad E\,\frac{d^2y}{dx^2} = \frac{2}{2}\beta;$$

for five observations

$$E\,\frac{dy}{dx} = \frac{1}{3}\beta,\quad E\,\frac{d^2y}{dx^2} = \frac{1}{3}\beta;$$

for seven observations

$$E\,\frac{dy}{dx} = \frac{1}{4}\beta,\quad E\,\frac{d^2y}{dx^2} = \frac{1}{6}\beta.$$

Thus we see that the errors in the first order derivatives decrease as do the inverse of the positive integers, and that the errors which affect the second order derivatives decrease according to the progression, even more rapid, of the inverse of the triangular numbers. The method of Mr. Laplace is therefore, from this point of view, entirely beyond reproach.

However, suppose we have between two fixed known limits sufficient observations to reduce to a very small value the difference between successive values of x. Following what we have just said, we must discard as many observations as we will use in our search for $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$. Thus, this is a serious inconvenience, especially if we have no reason to suspect that the values we discard are worse than those we use. In this manner we deprive ourselves of the compensation of errors upon which we may rely if we use all the values.

While reflecting on this subject, it seemed to me that it was possible, using the method of least squares,* to reconcile things and to obtain by this method all the precision one can possibly hope for in this situation. Here is the method I believe we should use.

Let a, a_1, a_2... be the values of x, however many, and let b, b_1, b_2,... be the observed corresponding values of y. Let

$$y = A + Bx + Cx^2 + Dx^3 + \ldots$$

allowing as many terms in this function as we would have employed using the previously described procedure of discarding observations. We wish to determine the value of the coefficients A, B, C, D,... . If the number of coefficients were equal to the number of observations, we could assign the coefficients values giving zero errors. But this is impossible in this case and we shall be content to minimize the sum of their squares.

These errors are respectively

$$A + Ba + Ca^2 + Da^3 + \ldots - b\ ;$$

$$A + Ba_1 + Ca_1^2 + Da_1^3 + \ldots - b_1;$$

$$A + Ba_2 + Ca_2^2 + Da_2^3 + \ldots - b_2.$$

We wish to obtain

$$\begin{aligned} &(A + Ba + Ca^2 + Da^3 + \ldots - b)^2 \\ +\ &(A + Ba_1 + Ca_1^2 + Da_1^3 + \ldots - b_1)^2 \quad = \text{minimum} \\ +\ &(A + Ba_2 + Ca_2^2 + Da_2^3 + \ldots - b_2)^2 \\ +\ &\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \end{aligned}$$

* We know that the method of least squares is based on the principle that the mean value (which is most probable to be nearly exact) of many values near a quantity, is that which, assuming it were correct, would minimize the sum of squares of the errors which affect the other observations. The first printed work in which this method was mentioned is the memoir of Mr. Legendre already cited in a preceding note (1806). In a work published in 1809, Mr. Gauss declared that he has been using a similar method since 1795. Mr. Laplace subsequently showed that this method conforms rigorously to the theory of probability.

That is, in differentiating with respect to A, B, C, D, ...

$$\begin{aligned}&(A + Ba + Ca^2 + \dots - b)(dA + adB + a^2dC + \dots)\\ &+(A + Ba_1 + Ca_1^2 + \dots - b_1)(dA + a_1dB + a_1^2dC + \dots) = 0\\ &+(A + Ba_2 + Ca_2^2 + \dots - b_2)(dA + a_2dB + a_2^2dC + \dots)\end{aligned}$$

Because of the independence between A, B, C, ... the multipliers of dA, dB, dC,... must separately be zero. We abbreviate in general

$$\Sigma a^m = a^m + a_1^m + a_2^m + \dots ;$$
$$\Sigma a^m b = a^m b + a_1^m b_1 + a_2^m b_2 + \dots ;$$

and we obtain these equations

$$\begin{aligned}&\Sigma a^0 A + \Sigma aB + \Sigma a^2 C + \Sigma a^3 D = \dots = \Sigma a^0 b,\\ &\Sigma aA + \Sigma a^2 B + \Sigma a^3 C + \Sigma a^4 D + \dots = \Sigma ab , \qquad (1)\\ &\Sigma a^2 A + \Sigma a^3 B + \Sigma a^4 C + \Sigma a^5 D + \dots = \Sigma a^2 b,\\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot\end{aligned}$$

There are exactly as many equations as there are unknown coefficients A, B, C, D,... . Although the methods previously discussed give values for y and its derivatives of a precision slightly inferior to that of the observations from which they were calculated, we can often hope with this new procedure to improve on the precision of the observations themselves.

The simplest case, and the most frequent, is that in which the values of x increase by a constant difference. Thus we can substitute the natural numbers for this progression. Let there be 2n + 1 known corresponding values of x and y. Let zero be the middle value of x, such that the numerical sequence is

$$-n, -(n-1), \dots -3, -2, -1, \pm 0, +1, +2, +3, \dots + (n-1), n.$$

Let Σn^m denote the sum of the m^{th} powers of these integers. We obtain

$$\Sigma a^0 = 2n + 1, \ \Sigma a = 0, \ \Sigma a^2 = 2\Sigma n^2, \ \Sigma a^3 = 0, \ \Sigma a^4 = 2\Sigma n^4, \ \dots .$$

Thus equations (1) become

$$(2n+1)A + 2\Sigma n^2C + \ldots = \Sigma b \quad , \quad 2\Sigma n^2B + 2\Sigma n^4D + \ldots = \Sigma ab \ ,$$

$$2\Sigma n^2A \quad + 2\Sigma n^4C + \ldots = \Sigma a^2b, \quad 2\Sigma n^4B + 2\Sigma n^6D + \ldots = \Sigma a^3b,$$

$$2\Sigma n^4A \quad + 2\Sigma n^6C + \ldots = \Sigma a^4b, \quad 2\Sigma n^6B + 2\Sigma n^8D + \ldots = \Sigma a^5b,$$

In addition to the fact that the sums of powers of the integers are given by known formulas, we also gain the advantage of being able to calculate separately the coefficients of even terms and those of odd terms, which will considerably simplify the amount of work.

Even in the case where neither the values of x nor the values of y occur in an arithmetical progression, we can profit from these simplifications by proceeding as follows. Suppose that x and y are both functions of a third variable z, whose values are completely arbitrary, but are equally spaced, as with x above. We would seek by our procedure the values of $\frac{dx}{dz}, \frac{dy}{dz}, \frac{d^2x}{dz^2}, \frac{d^2y}{dz^2}, \ldots$. We would then obtain, using known formulas,

$$\frac{dy}{dx} = \frac{dy/dz}{\frac{dx}{dz}}, \quad \frac{d^2y}{dx^2} = \frac{\frac{dx}{dz}\frac{d^2y}{dz^2} - \frac{dy}{dz}\frac{d^2x}{dz^2}}{\left(\frac{dx}{dz}\right)^2}$$

This method seems to me preferable to that which consists of interpolation between observations in order to render them equidistant. It is understood, of course, that it may be dangerous, in a problem of a rather delicate nature, to change the values of the observations before using them.

It seems to us that the introduction of the method which we have described into the method of Mr. Laplace, for the determination of the orbit of comets, will greatly increase its precision, at least in the case where we have a large number of observations. However, this method, as is true of many other methods, will basically be nothing more than well-directed groping.

There remains another problem to be resolved, which can be stated as follows: *we know that a number of points, however many, are located near a parabolic curve of unknown fixed degree, and we wish to know the most likely value of the degree of this curve.* The solution to this problem would eliminate the uncertainty of the analyst who, wishing to apply the method of Mr. Laplace, is able to employ a large number of observations.