

ART. III.—*On the Empirical Interpolation of Observations in Physics and Chemistry*; by W. P. G. BARTLETT.

THE object of the present paper is to bring to the notice of physicists some methods of interpolation; not that there is any principle in them new to mathematicians, but because no proper methods appear to be practically within the reach of many of those engaged in making such observations in physics and chemistry as require interpolation.

Whatever difficulty there is in the problem arises from our entire ignorance of the form of the function which the observations follow, and from the necessarily irregular intervals at which they are made.

Every method of interpolation under these conditions amounts to *assuming* some formula involving arbitrary constants, and determining the values of these by elimination from the equations furnished by the observations.

If all the observations are required to be rigorously satisfied, Lagrange's formula,

$$y = \frac{(t-t_1)(t-t_2)(t-t_3)\dots}{(t_0-t_1)(t_0-t_2)(t_0-t_3)\dots} y_0 + \frac{(t-t_0)(t-t_2)(t-t_3)\dots}{(t_1-t_0)(t_1-t_2)(t_1-t_3)\dots} y_1 \\ + \frac{(t-t_0)(t-t_1)(t-t_3)\dots}{(t_2-t_0)(t_2-t_1)(t_2-t_3)\dots} y_2 + \&c.,$$

is probably as good a way as any of arranging the elimination, since it is only necessary to multiply the various factors to obtain the function y , developed in powers of the variable t ; y_0 , y_1 , &c., being the observed values corresponding to the values t_0 , t_1 , &c., of t^* . Otherwise the determination of the constants may be made either so as to satisfy exactly some of the observations, or so as to satisfy them all within moderate limits—say the probable errors of the observations themselves. The former proceeding is theoretically imperfect, because it makes some of the observations of no account whatever in determining the values of the constants, using them, if at all, only to help the selection by successive trials of the form of the function. The latter is generally impracticable in a direct form, unless the constants enter linearly into the equations, in which case the method of *least squares* will always give good results; but if besides this, the successive terms in the development, either of y , or of any given function of y , form a convergent series, it will generally be advantageous to use Cauchy's method, which, *notwithstanding its violation of the law of probable error*, is practically sufficient, and indeed far the best, for almost all the physical formulæ that it is worth while to develop at all in an *empirical series*.

This method not being, like *least squares*, generally accessible in a working form, it is proposed to devote special attention to its operation. Its principle is to neglect at each step all the terms of lower orders, leaving in general a form

$$z = a u,$$

and then of all the values

$$\frac{k_0 z_0 + k_1 z_1 + k_2 z_2 + \&c.}{k_0 u_0 + k_1 u_1 + k_2 u_2 + \&c.}$$

which might be given to the constant a , by assigning different sets of values to the k 's, to select that in which the k 's are all so taken ($=\pm 1$) that the denominator above written becomes the

* If there are $m+1$ observed values of y , there will be only $\frac{1}{2}m(m+1)$ different factors to be computed in all the denominators of the formula above written; and it is evident from the theory of equations, that the coefficient of t^{m-r} in the numerator (expanded in powers of t) of the coefficient of y_n will be $(-1)^r$ times the sum of the products formed by every possible combination of r different factors, t_0 , t_1 , &c., omitting from the sum the terms containing t_n ; which involves for the whole work only $2^{m+1} - (m+3)$ different products of two or more terms each, and therefore only this number of multiplications.

absolute sum of the special values of u .* To show more distinctly how the numerical application is to be made, we shall here arrange some formulæ for computation, changing for this purpose some of Cauchy's notation and giving the development an entirely different form.

Let it be assumed, as usual, that the observed quantity y , converges when developed in the form

$$(1) \quad y = A + Bt + Ct^2 + \&c.$$

If this assumption does not give, on trial, a convenient formula, the logarithm or any other function of y may be tried in the place of y , and its development made in precisely the same way. It will be easy to see, moreover, that any variables we please may be substituted for the different powers of t , *provided only the series is convergent*. The function will first be developed in the form

$$(2) \quad y = \mathfrak{A} + \mathfrak{B}\Delta t + \mathfrak{C}\Delta^2 t^2 + \mathfrak{D}\Delta^3 t^3 + \&c.$$

in which Δt , $\Delta^2 t^2$, &c., are functions of the form $a + bt + ct^2 + \&c$. and are respectively of the first, second, &c., degrees in t . The numerical values of \mathfrak{A} , \mathfrak{B} , &c., and the expressions for Δt , &c., being found, the series (2) is immediately reducible to the form (1). Let s be the number of observations given to determine A , B , &c.; then the formulæ required in practice are

$$(3) \quad \left\{ \begin{array}{l} \alpha_1 = \frac{\Sigma t}{s}, \quad \alpha_2 = \frac{\Sigma t^2}{s}, \quad \alpha_3 = \frac{\Sigma t^3}{s}, \dots \dots \dots \alpha_n = \frac{\Sigma t^n}{s} \\ \Delta t = t - \alpha_1, \quad \Delta t^2 = t^2 - \alpha_2, \quad \Delta t^3 = t^3 - \alpha_3, \dots \dots \Delta t^n = t^n - \alpha_n \\ \beta_2 = \frac{\Sigma' \Delta t^2}{\Sigma' \Delta t}, \quad \beta_3 = \frac{\Sigma' \Delta t^3}{\Sigma' \Delta t}, \dots \dots \dots \beta_n = \frac{\Sigma' \Delta t^n}{\Sigma' \Delta t} \\ \Delta^2 t^2 = \Delta t^2 - \beta_2 \Delta t, \quad \Delta^2 t^3 = \Delta t^3 - \beta_3 \Delta t, \dots \Delta^2 t^n = \Delta t^n - \beta_n \Delta t \\ \gamma_3 = \frac{\Sigma'' \Delta^2 t^3}{\Sigma'' \Delta^2 t^2}, \dots \dots \dots \gamma_n = \frac{\Sigma'' \Delta^2 t^n}{\Sigma'' \Delta^2 t^2} \\ \Delta^3 t^3 = \Delta^2 t^3 - \gamma_3 \Delta^2 t^2 \dots \Delta^3 t^n = \Delta^2 t^n - \gamma_n \Delta^2 t^2 \\ \mathfrak{A} = \frac{\Sigma y}{s} \quad y' = y - \mathfrak{A} \\ \mathfrak{B} = \frac{\Sigma' y'}{\Sigma' \Delta t} \quad y'' = y' - \mathfrak{B} \Delta t \\ \mathfrak{C} = \frac{\Sigma'' y''}{\Sigma'' \Delta^2 t^2} \quad y''' = y'' - \mathfrak{C} \Delta^2 t^2 \\ \mathfrak{D} = \frac{\Sigma''' y'''}{\Sigma''' \Delta^3 t^3} \quad y^{iv} = y''' - \mathfrak{D} \Delta^3 t^3 \end{array} \right.$$

* For the complete analysis, which is quite simple, the reader is referred to the original lithographed memoir published in 1835, or to its republication in 1837 in Liouville's *Journal de Mathématiques*, tome ii, page 198. The same thing is also appended as a note to the first volume of Moigno's *Calcul Différentiel*, page 513; and a partial translation of it in the *U. S. Coast Survey Report* for 1860, p. 392.

Σ , Σ' , &c., indicate the algebraic sum of the s values of the respective functions before which they are placed; but before taking the sum Σ' , the signs of all the numbers corresponding to the cases in which the values of Δt are negative, and including these values themselves, must be changed. Similarly before taking Σ'' , Σ''' , &c., signs must be changed throughout for the cases in which $\Delta^2 t^2$, $\Delta^3 t^3$, &c., respectively are negative. So that $\Sigma' \Delta t$, $\Sigma'' \Delta^2 t^2$, $\dots \Sigma^{[m]} \Delta^m t^m$ are the absolute sums of these quantities. The following equations

$$(4) \quad \Sigma^{[m]} \Delta^n t^n = 0, \quad \Sigma^{[m]} y^{[n]} = 0$$

are true for all values of n greater than m , and may therefore be used as *checks*. Each of the conditions (4) breaks up (except the case in which $n=1$) into two more convenient partial sums; for, denoting the sum of all the values of a function corresponding to positive values of $\Delta^m t^m$ by $\Sigma^{[m]}(+)$, and of those corresponding to negative ones by $\Sigma^{[m]}(-)$, the equation

$$\Sigma = 0 \text{ is equivalent to } \Sigma^{[m]}(+) + \Sigma^{[m]}(-) = 0, \text{ and}$$

$$\Sigma^{[m]} = 0 \quad \text{ " } \quad \Sigma^{[m]}(+) - \Sigma^{[m]}(-) = 0; \text{ whence}$$

$$(5) \quad \Sigma^{[m]}(+) = 0, \quad \Sigma^{[m]}(-) = 0,$$

which may take the places of Σ and $\Sigma^{[m]}$ in the form (4). There might occur cases in which this principle of subdivision could be carried on still farther. The advantage of using (5) instead of (4) lies in the narrower limits within which it is necessary to look for an error discovered by means of (5).

The special forms of the various functions are written out in (3) as far as will suffice for determining four terms in the value of y , and computing y^{iv} so as to test the accuracy of the approximation and apply the checks (5) to it. An inspection of (3) will show:

1st, that the first term, \mathfrak{A} , is simply the *average value* of y .

2d, that to determine the second term it will be necessary to compute y' , α_1 , Δt , and \mathfrak{B} :

3d, for the third term, y'' , α_2 , Δt^2 , β_2 , $\Delta^2 t^2$, and \mathfrak{C} :

4th, for the fourth term, y''' , α_3 , Δt^3 , β_3 , $\Delta^2 t^3$, γ_3 , $\Delta^3 t^3$, and \mathfrak{D} : and so on till the residual quantities, $y^{[m]}$, are seen to be small enough to be neglected.

If more special forms are desired besides those written out in (3), the law of their formation is obvious from an inspection of those actually developed there. It is such that, in general, if μ and \mathfrak{M} be the m th letters in their respective alphabets, then

$$\mu_n = \frac{\Sigma^{[m-1]} \Delta^{m-1} t^n}{\Sigma^{[m-1]} \Delta^{m-1} t^{m-1}}, \quad \Delta^m t^n = \Delta^{m-1} t^n - \mu_n \Delta^{m-1} t^{m-1}$$

$$\mathfrak{M} = \frac{\Sigma^{[m-1]} y^{[m-1]}}{\Sigma^{[m-1]} \Delta^{m-1} t^{m-1}}, \quad y^{[m]} = y^{[m-1]} - \mathfrak{M} \Delta^{m-1} t^{m-1}.$$

It will be observed that no cases occur in which n is less than m .

To get out symmetrically the coefficients of (1) it is easy to find that they are of the following forms:

$$(6) \quad \begin{cases} A = \mathfrak{A} + [1]\mathfrak{B} + [2]\mathfrak{C} + [3]\mathfrak{D} + \&c. \\ B = \mathfrak{B} + [2']\mathfrak{C} + [3']\mathfrak{D} + \&c. \\ C = \mathfrak{C} + [3'']\mathfrak{D} + \&c. \\ D = \mathfrak{D} + \&c. \end{cases} \quad \begin{matrix} \text{etc.} & \text{etc.} & \text{etc.} \end{matrix}$$

in which

$$\begin{aligned} [1] &= -\alpha_1 \\ [2] &= -\alpha_2 - \beta_2[1] & [2'] &= -\beta_2 \\ [3] &= -\alpha_3 - \beta_3[1] - \gamma_3[2] & [3'] &= -\beta_3 - \gamma_3[2] & [3''] &= -\gamma_3 \\ &\text{etc.} & & \text{etc.} & & \text{etc.} \end{aligned}$$

EXAMPLE.

[Löwel's Solubility of Anhydrous NaO SO₃ in water, developed in powers of $t=50^\circ$.—*Annales de Chimie et Physique*, xlix, 50.]

	y	t	t^2	t^3	y'	Δt	Δt^2	Δt^3
1.	5325	-32.00	1024.0	-32768	+487.5	-27.684	+400.9	-39445
2.	5276	30.00	900.0	27000	438.5	25.684	276.9	33677
3.	5153	25.00	625.0	15625	315.5	20.684	+1.9	22302
4.	5131	24.00	576.0	13824	293.5	19.684	-47.1	20501
5.	5037	20.00	400.0	8000	199.5	15.684	223.1	14677
6.	4971	17.00	289.0	4913	133.5	12.684	334.1	11590
7.	4953	16.00	256.0	4096	115.5	11.684	367.1	10773
8.	4878	9.85	97.0	956	+40.5	5.534	526.1	7633
9.	4781	-4.96	24.6	-122	-56.5	-0.644	598.5	6799
10.	4632	+0.40	0.2	0	155.5	+4.716	622.9	6677
11.	4542	9.79	95.8	+938	295.5	14.106	527.3	-5739
12.	4435	20.61	424.8	8755	402.5	24.926	-198.3	+2078
13.	4296	34.42	1184.7	40779	541.5	38.736	+561.6	34102
14.	4265	+53.17	2827.0	+150314	-572.5	+57.486	+2203.9	+143637
Σ	67725	-60.42	8724.1	+93482	0	+0.004	+0.7	+4
Σ'					-3935.0	279.936	+2833.3	+334798

	y''	$\Delta^2 t^2$	$\Delta^2 t^3$	y'''	$\Delta^3 t^3$	y^{iv}	
1.	+98.4	+681.2	-6385	+6.3	-18220	+8.3	$\alpha_1 = -4.316$
2.	77.5	536.9	-2959	+4.9	12326	+6.3	$\alpha_2 = +623.1$
3.	24.8	211.3	+2436	-3.8	-1251	-3.7	$\alpha_3 = +6677$
4.	+16.8	+152.2	3041	3.8	+386	3.8	
5.	-21.0	-64.3	4081	12.3	5203	12.9	$\beta_2 = +10.125$
6.	44.8	205.7	3580	17.0	7169	17.8	$\beta_3 = +1196.0$
7.	48.7	248.8	+3201	-15.1	7542	-15.9	
8.	37.3	470.1	-1014	+26.3	7188	+25.5	$\gamma_3 = +17.447$
9.	65.6	592.0	6029	14.3	+4300	13.8	
10.	89.2	670.6	12317	+1.5	-617	+1.6	$\mathfrak{A} = 4837.5$
11.	97.2	670.1	22610	-6.6	10916	-5.4	
12.	-52.1	-450.7	27733	+8.8	19870	+11.0	$\mathfrak{B} = -14.057$
13.	+3.0	+169.4	-12227	-19.9	-15182	-18.2	
14.	+231.5	+1621.9	+74884	+16.2	+46587	+11.0	$\mathfrak{C} = +0.1352$
$\Sigma'(+)$	0	-0.1	-3	0	+2	0	
$\Sigma'(-)$	+0.1	+0.7	+2	-2	-9	-2	$\mathfrak{D} = +0.000111$
Σ''	+911.9	6745.2	+117681	-1.1	-6.1	-1.1	
Σ'''				+17.4	156757	-1.1	

	$\beta_2 \Delta t$	$\beta_3 \Delta t$	$\beta_4 \Delta t$	$\gamma_3 \Delta^2 t^2$	$\mathcal{C} \Delta^2 t^2$	$\mathcal{B} \Delta^3 t^3$	
1.	- 280.3	- 33110	+ 389.1	+ 11885	+ 92.1	- 2.0	[1] = + 4.816
2.	260.0	30718	361.0	9367	72.6	1.4	[2] = - 666.8
3.	209.4	24738	290.7	3687	28.6	- 0.1	[3] = - 206
4.	199.3	23542	276.7	+ 2655	+ 20.6	0.0	
5.	158.8	18758	220.5	- 1122	- 8.7	+ 0.6	[2'] = - 10.125
6.	128.4	15170	178.3	3589	27.8	0.8	[3'] = - 1019
7.	118.3	13974	164.2	4341	33.6	0.8	
8.	56.0	6619	77.8	8202	63.6	0.8	[3''] = - 17.4
9.	- 6.5	- 770	+ 9.1	10328	79.9	+ 0.5	
10.	+ 47.7	+ 5640	- 66.3	11700	90.7	- 0.1	A = 4686.6
11.	142.8	16871	198.3	11694	90.6	1.2	B = - 15.539
12.	252.4	29811	350.4	- 7863	- 60.9	2.2	C = + 0.1333
13.	392.2	46329	544.5	+ 2955	+ 22.9	- 1.7	D = + 0.000111
14.	+ 582.0	+ 68753	- 808.0	+ 28297	+ 219.3	+ 5.2	

In this example, to which Cauchy's method is applied, the numbers in the column y express the solubility* of anhydrous sulphate of soda in 10,000 parts of water at 14 different temperatures,—column t contains the corresponding temperatures less 50° (for convenience of development). The work is carried as far as the determination of four terms in the development according to powers of (temp. -50°); but a comparison of the values of y^{iv} with those of y''' shows that nothing is gained by the addition of the fourth term. Since the values of A, B, &c., are obtained by using (6) it is not necessary to obtain the coefficients a , b , &c., at all. In taking the sum Σ' the signs are to be changed in the first nine cases; therefore $\Sigma'(+)$ means the algebraic sum of the last five cases, and $\Sigma'(-)$ that of the first nine. $\Sigma''(+)$ is the algebraic sum of the 1st, 2d, 3d, 4th, 13th and 14th cases, and $\Sigma''(-)$ of the rest. $\Sigma'''(+)$ includes the 4th, 5th, 6th, 7th, 8th, 9th, and 14th, and $\Sigma'''(-)$ the rest. The values of $\Sigma''(+)$ and $\Sigma''(-)$ are written on the same horizontal line opposite the argument Σ'' ; similarly with $\Sigma'''(+)$ and $\Sigma'''(-)$ for y^{iv} .

M. Bienaymé has shown† that if each case of the equations,

$$\begin{aligned}
 y &= A + Bt + Ct^2 + \&c. \\
 y' &= B + 2Ct + \&c. \\
 y'' &= 2C + 2Dt + \&c. \\
 \text{etc.} & \qquad \qquad \text{etc.} \qquad \qquad \text{etc.}
 \end{aligned}$$

were multiplied by the proper *least-square* factor (different for each case and for each of these equations) before taking the sums Σ , Σ' , &c., the process would become merely another form for the expression of the elimination in *least-squares* given by Gauss‡ and others.

The advantages of Cauchy's method are its simplicity, the ease

* According to Löwel, *Annales de Chimie and Physique*, xlix, page 50.

† *Comptes Rendus*, tome xxxvii, 4 Juillet 1853; and Liouville's *Journal de Mathématiques*, xviii, page 299. See also Cauchy's Note, *Comptes Rendus*, xxxvi, 27 Juin, 1853.

‡ In the *Disquisitio de elementis Palladis*; and translated into French on page 137 of the *Méthode des Moindres Carrés*.

with which the important check (5) is applied to the work, and the fact, which is of great importance in many of its applications to physics, that it is not necessary to determine beforehand how many coefficients A , B , &c., are to be eliminated. The objections, which seem fatal to it as a substitute for *least squares* where the latter is properly applicable, are of very little importance in cases where the form of the function is wholly assumed and the formula therefore only to be trusted within the limits of the series of observations; for in these cases the formula which gives a minimum value to the sums of the squares of the differences between the computed and observed quantities, is not necessarily better than many others giving other sets of properly distributed small differences, especially as "the errors of observation" are frequently mixed up with others of the same order of magnitude, arising from errors in the values of the variables, t , t^2 , &c.

If however the series of observations is very extensive (like those, for instance, on the tension of steam) the labor of finding an empirical formula becomes altogether greater than its value, and it is better to *tabulate* the function without reference to any "interpolation-formula" satisfying the whole or even any great number of the observations. For this purpose the observed quantities must first be reduced to equidistant values of the variable, and then these may easily be interpolated to as frequent intervals as we please by the methods in common use.

To accomplish the first object there are several methods. The mechanical one of *plotting a curve* (however valuable in suggesting the true physical law of the phenomena) cannot often be used for this purpose with as much accuracy as computation. In the method of interpolation by "*divided differences*"* each determined place depends only on a very few of the adjacent observations, and a series of such places, unless the observations were accurate to the last figure, would not be apt to harmonize. Another way is as follows: let one of the equidistant values of the variables be t_0 , then the observations may be represented, in the vicinity of t_0 , by the series,

$$(7) \quad y = A + B(t - t_0) + C(t - t_0)^2 + \&c.$$

in which A is obviously the required value of y corresponding to $t = t_0$, and may be determined in each case from as many observations as we please to use. Cauchy's method applied in this way to some of Regnault's observations has been found to give an accurate table of vapor-tensions with very little labor. Determinations of A were made for every 6° of temperature from 9° to 39° , and ten observations were used for each determination.

* See Lacroix, *Calcul*, tom. iii, p. 31, § 903; or De Morgan, *Calculus*, p. 550.

AM. JOUR. SCI.—SECOND SERIES, VOL. XXXIV, No. 100.—JULY, 1863.