Sur les différences qui distinguent l'interpolation de M. Cauchy de la méthode des moindres carrés, et qui assurent la supriorité de cette méthode*

M. Jules Bienaymé

Comptes Rendus de l'Académie des Sciences XXXVII, pp. 5–13
Liouville's Journal Math. Pures et Appliques 18, pp. 299–308

Session of 4 July 1853

Since some time, the attention of many observers is carried on a *method of interpolation* that Mr. Cauchy has published in 1835,[1] and it seems that one has regarded this method as having something of analogy to the advantages of the celebrated method of least squares. It would be annoying that the observers were deceived in this regard by that which has been able to be said of the two methods, because they differ completely; and if the process of Mr. Cauchy bears witness, as all that which exits from his pen, of the ingenious industry that he knows how to bring as far as into practical questions, this process is no less completely in contradiction with the principles of the calculation of probabilities. This discord does not appear to be known, although it is very easy to perceive it. But it is this that the limited time of which the observers make the sacrifice to analysis, does not permit them to research. A warning is able to be useful to them; and without touching in the least to the value of each will link to the process of Mr. Cauchy, as a means of interpolation (of convergent series especially), it will be permitted to show that this process is only a modification of ordinary elimination among many equations of the first degree; a modification already prescribed by the authors who are occupied with least squares, and that Mr. Gauss has reduced to an algorithm; that it offers no special degree of probability when one applies it to some equations more numerous than the unknowns to determine; that on the contrary, it adds then to the risks of error, and it does not assign to it the measure; finally, that, as means of elimination, it would be applied perfectly to the method of least squares, if by chance the number of unknowns were too considerable in order that one wished to calculate them all, and that the last gave besides only some terms less than the errors of the quantities observed, supposed only in a member of the equation. It is true that then nothing would be more simple than to suppress beforehand these unknowns, that

*Translated by Richard J. Pulskamp, Department of Mathematics & Computer Science, Xavier University, Cincinnati, OH. June 19, 2010

[1] "Mémoire sur l'interpolation." Liouville's Journal II, 1837, p. 193-205.

would cause to recognize the previous examination which must always be made of the equations considered, before applying any process.

In order to show, as simply as the matter includes it, how the process of M. Cauchy is only a modification of ordinary elimination, it suffices to resume this elimination. Let therefore there be a number n of equations, among all as many unknowns, of the form

$$(1) x_1 a_h + x_2 b_h + x_3 c_h + \cdots + x_n l_h = w_h.$$

If one multiplies all these equations by a first system of arbitrary factors $k_{1,h}$, and if one adds the products; next by a second system of factors $k_{2,h}$, and if one adds likewise the products; and if one repeats this operation n times, it is clear that one will obtain n equations of the form

$$(2) \qquad x_1 S.a_h k_{1,h} + x_2 S.b_h k_{1,h} + x_3 S.c_h k_{1,h} + \cdots + x_n S.l_h k_{1,h} = S.w_h k_{1,h};$$

$$(3) \qquad x_1 S.a_h k_{i,h} + x_2 S.b_h k_{i,h} + x_3 S.c_h k_{i,h} + \cdots + x_n S.l_h k_{i,h} = S.w_h k_{i,h}.$$

Whatever be the arbitrary factors k, these equations will be able to replace the first, provided that each system of factors is different, finally that the new equations do not reenter the ones into the others. But it is palpable that the choice of these factors will not influence on the final values of the unknowns, of which they will vanish entirely when there will be as many unknowns as originally given equations. It would not be likewise if there were more equations than unknowns; but it is a point on which it will suffice to return later, because nothing prevents, in this same case, to suppose first as many unknowns as one will wish, except to annul next the coefficients of a part of these unknowns.

Now one will proceed to the elimination of x_1 among the first of the new equations and each of the $(n-1)$ others, as usually, by rendering equal the coefficients of this unknown, and subtracting successively the first equation from each of the $(n-1)$ others.

For example, in order to subtract the equation of rank i, which has been written above (3) as model of all, one multiplies the first by the ratio of the coefficients of x_1,

$$\frac{S.a_h k_{i,h}}{S.a_h k_{1,h}},$$

and one obtains without pain an equation no longer containing x_1:

$$\begin{aligned} & x_2 \left(S.b_h k_{i.h} - S.b_h k_{i,h} \frac{S.a_h k_{i,h}}{S.a_h k_{1,h}} \right) \\ & + x_3 \left(S.c_h k_{i.h} - S.c_h k_{i,h} \frac{S.a_h k_{i,h}}{S.a_h k_{1,h}} \right) + \cdots \\ & + x_n \left(S.l_h k_{i.h} - S.l_h k_{i,h} \frac{S.a_h k_{i,h}}{S.a_h k_{1,h}} \right) \\ & = S.w_h k_{i,h} - S.w_h k_{i,h} \frac{S.a_h k_{i,h}}{S.a_h k_{1,h}}. \end{aligned}$$

There will be $(n-1)$ equations of this form, and they will contain no more than $(n-1)$ unknowns.

If one pays attention to the composition of the coefficients of these equations, one sees that any one

$$S.c_h k_{i.h} - S.c_h k_{1,h} \frac{S.a_h k_{i,h}}{S.a_h k_{1,h}}$$

is able to be written:

$$\begin{aligned} &S.c_h k_{i.h} - S.c_h k_{1,h} \frac{S.a_h k_{i,h}}{S.a_h k_{1,h}} \\ =&S.k_{i.h} \left(c_h - a_h \frac{S.c_h k_{1,h}}{S.a_h k_{1,h}} \right) \\ =&S, k_{i,h} \Delta c_h, \end{aligned}$$

provided that one has named, with Mr. Cauchy, by Δc_h, the differences between parentheses, and that one has formed all these differences. These will be

$$\begin{aligned} b_h - a_h \frac{S.b_h k_{1,h}}{S.a_h k_{1,h}} &= \Delta b_h, \\ c_h - a_h \frac{S.b_h k_{1,h}}{S.a_h k_{1,h}} &= \Delta c_h, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ w_h - a_h \frac{S.w_h k_{1,h}}{S.a_h k_{1,h}} &= \Delta w_h, \end{aligned}$$

Then the $(n-1)$ equations, among the $(n-1)$ unknowns, takes the form

$$\begin{aligned} x_2 S.\Delta b_h k_{2,h} + x_3 S.\Delta c_h k_{2,h} + \cdots = S.\Delta w_h k_{2,h}, \\ x_2 S.\Delta b_h k_{i,h} + x_3 S.\Delta c_h k_{i,h} + \cdots = S.\Delta w_h k_{i,h}. \end{aligned}$$

It is clear that one will form these equations, by subtracting the sum (2) of the products by the factors $k_{i,h}$ of the n given equations (1), from each of these equations, after having multiplied this sum by $\frac{a_h}{S.a_h k_{1,h}}$. There would come thus $(n-1)$ equations of the form

$$x_2 \Delta b_h + x_3 \Delta c_h + \cdots + x_n \Delta l_h = \Delta w_h;$$

and by multiplying them by the factors of the system $k_{2,h}$, one will fall again onto the first of the $(n-1)$ equations already obtained. The others would depend on the system of factors designated by $k_{i,h}$, etc.

Nothing comes therefore to mix the second system of factors with the first; and it is put likewise that if one had introduced it only after the elimination of the first unknown x_1; but, as one has seen, it is absolutely as if one had introduced all first.

At present, nothing is more easy than to pursue the elimination of the unknowns one after the other. In representing by Δ^2 the differences formed with the differences Δ and the system of factors $k_{2,h}$, as the Δ has been with the coefficients and the factors $k_{1,h}$; for example,

$$\begin{aligned} \Delta^2 c_h &= \Delta c_h - \Delta b_h \frac{S.\Delta c_h k_{2,h}}{S.\Delta b_h k_{2,h}}, \\ \Delta^2 w_h &= \Delta w_h - \Delta b_h \frac{S.\Delta w_h k_{2,h}}{S.\Delta b_h k_{2,h}}, \end{aligned}$$

it is clear that one will arrive to $(n-2)$ equations among $(n-2)$ unknowns, of the form

$$x_3 S.\Delta^2 c_h k_{3,h} + \cdots + S.\Delta^2 l_h k_{3,h} = S.\Delta^2 w_h k_{3,h};$$
$$x_3 S.\Delta^2 c_h k_{i,h} + \cdots + x_n S..\Delta^2 l_h k_{i,h} = S..\Delta^2 w_h k_{i,h}.$$

In going next in the same manner concerning these equations, one would eliminate next an unknown; and it is superfluous to push further the operation. The end is actually reached: it was to show that of the factors $k_{i,h}$, whatever they be (Mr. Cauchy, one knows, takes them equal to ± 1), can be introduced from the beginning of the operation, without modifying in the least the results. One was able to think that Mr. Cauchy introducing the second system of factors only after having formed the differences Δ, this system would be to incur some special condition if one would wish to reascend to the combination of the original equations which, based on this second system, would leave however each $k_{2,h}$ completely arbitrary. One was able to fear that $k_{3,h}$ not come to require some factors complicated by the operations which lead to the successive equations. It is easy at present to recognize that the things are not so complicated, and that the successive elimination of one unknown leaves outside of the calculations all the systems of factors of a higher index than the index of this unknown. So that these factors enjoy the same role as if they came to be introduced arbitrarily.

Now they give the same results as equations (2) and (3), where they are introduced since the origin. And it is very easy to see that they give the same results, to any number $m < n$ that one reduces these equations, and the unknowns that they contain. Because one will arrive successively (and it is there the ingenious side of the process, that one attributes to Mr. Cauchy or to Mr. Gauss[2]), one will arrive to n equations:

$$\begin{aligned}
x_1 S.a_h k_{1,h} + x_2 S.b_h k_{1,h} + x_3 S.c_h k_{1,h} + \cdots + x_n S.l_h k_{1,h} &= S.w_h k_{1,h},\\
x_2 S.\Delta b_h k_{2,h} + x_2 S.\Delta c_h k_{2,h} + \cdots + x_n S.\Delta l_h k_{2,h} &= S.\Delta w_h k_{2,h},\\
x_3 S.\Delta^2 c_h k_{3,h} + \cdots + x_n S.\Delta^2 l_h k_{3,h} &= S.\Delta^2 w_h k_{3,h},\\
&\vdots\\
x_n S.\Delta^{n-1} l_h k_{n,h} &= S.\Delta^{n-1} w_h k_{n,h}.
\end{aligned}$$

That if one is arrested at m unknowns, the first terms, in which these unknowns enter, will be precisely the same as if one had taken the entire equations.

On another side, one sees very clearly that these first terms, deduced from n equations of the form

$$x_1 a_h + x_2 b_h + \cdots x_n g_h = w_h,$$

present the ordinary result of elimination among these equations reduced to the number m, by the multiplication of m systems of n arbitrary factors and by the addition of the products.

Now one knows, by the method of least squares, what these m systems each of n factors must be, in order that the final error due to the partial errors of the observed quantities w_h be a minimum; in other terms, in order that the result be such, that the

[2]That which distinguishes especially the process of Mr. Cauchy, is the calculation of the remainders $\Delta^i w_h$, at each elimination of the unknowns.

sum of the squares of the differences among the w_h and the first members of the n given equations becomes a minimum. The factors $k_{i,h}$, in order to satisfy this condition evidently advantageous, must be the same coefficients of the unknowns. The factors of Mr. Cauchy are, on the contrary, all equal to ± 1: they could therefore give a result neither as probable nor as advantageous as is the one of the method of least squares.

There is more: these factors assign to the result no special probability; because they take their signs in a manner that always $\Delta^{i-1} g_h k_{i,h}$ is positive, if g_h indicates precisely the coefficients of the unknown x_i of rank i, with which the factors $k_{i,h}$ appear, in the order of elimination according to Mr. Cauchy. Now there is nothing there which assigns rather a magnitude than another to the errors that one leaves to subsist.

Let one consider effectively the first result

$$\frac{S.w_h k_{1,h}}{S.a_h k_{1,h}}.$$

The factors $k_{i,h}$ are ± 1, taken in a manner that $S.a_h k_{1,h}$ is equal to the sum of the absolute values of a_h. Hence

$$\frac{S.w_h k_{1,h}}{S.a_h k_{1,h}}.$$

is an entire mean between the greatest and the smallest of the fractions with positive denominator,

$$\frac{w_h}{a_h}.$$

If ϵ_h is the error of w_h, the error of

$$\frac{S.w_h k_{1,h}}{S.a_h k_{1,h}} \quad \text{will be} \quad \frac{S.\epsilon_h k_{1,h}}{S.a_h k_{1,h}},$$

that is to say a mean between the fractions $\frac{\epsilon_h}{a_h}$. There results from it that the error of

$$\Delta w_h = w_h - a_h \frac{S.w_h k_{1,h}}{S.a_h k_{1,h}}$$

will be

$$\epsilon_h - a_h \frac{S.\epsilon_h k_{1,h}}{S.a_h k_{1,h}};$$

and as $\frac{S.\epsilon_h k_{1,h}}{S.a_h k_{1,h}}$ is all the more equal to $\frac{\epsilon_u}{a_u}$, the greatest of the fractions $\frac{\epsilon_h}{a_h}$, one will have only

$$\epsilon_h - a_h \frac{\epsilon_u}{a_u}.$$

Now, because of the sign and of the magnitude of a_h relatively to the absolute value of a_u, it will be able that $\epsilon_h - a_h \frac{\epsilon_u}{a_u}$ makes a sum superior to the greatest of the errors ϵ_h.

That which comes to be said is applied to all the degrees of operation, so that nothing guarantees that the errors will not be increasing.

But this is not all: if the operation is arrested at any one unknown, it introduces by its same nature another kind of error; since one neglects then one sequence of terms which, in each equation, must not be neglected, but subtracted from w_h before proceeding with the elimination. Calling these neglected quantities δ_h, it is clear that it will arrive to δ_h that which one comes to recognize for the ϵ_h; and that the combinations $\Delta\delta_h$, $\Delta^2\delta_h$, $\Delta^3\delta_h$, etc., will be able to grow and not decrease in the sequence of calculations. It will be quite difficult to be averted from it; because the quantities Δw_h, $\Delta^2 w_h$, etc., that Mr. Cauchy takes for indices of the term of the operation, are subject themselves to increase and to decrease. One sees an example in the interpolation even made by the author, and published in the *Nouveaux Exercices de Mathématiques*, Prague, 1835. Thus one is not sure that it is necessary to arrest the calculations according the magnitude of these indices.

It is necessary to hurry to add that Mr. Cauchy has proposed his method only in order to interpolate some series of which the convergence is assured previously; and that under this particular circumstance, the Δw_h, $\Delta^2 w_h$, etc., will go without doubt by diminishing. But his example even proves that this special case is not exempt from the difficulty signaled; and yet the convergence was very great. It is quite clear that this difficulty will affect quite so much the more the use that one will be able to make of his method to some equations of condition, where the unknowns and their coefficients do not form a very convergent series in the first member. Now one seems today to wish to make of this method a general rule, equally good in all cases.

One sees that this is not; that this is uniquely a way of elimination which is able to offer advantages under certain circumstances. Interpolation is a problem so indeterminate, that it is good to have diverse processes, even in order to eliminate among the equations to which on decides that it is necessary to stop. To this claim, it will be to the observer to discuss the problem that it is necessary to resolve; and to establish if there is for it some utility to apply the process to Mr. Cauchy, instead of the methods of interpolation that one employs more often. But when he would wish to obtain the minimum errors, one sees that he should not substitute this process for the method of least squares.

Moreover, after all that which precedes, it is clear that the coefficients $k_{i,h}$ are able to be those of the method of least squares. It is therefore very practical, under this method, to make the successive eliminations, by transforming the system of equations in m unknowns, into a system which will have only one equation in m unknowns, one in $(m-1)$ unknowns, one in $(m-2)$, and thus in sequence, until one equation in one unknown alone, and, if one wishes, to take into consideration the magnitude of the remainders.

If therefore some particular advantage is encountered in the process, one will obtain without sacrifice in the least the advantages well superior to the method of least squares. Thus had Laplace prescribed precisely the same mode of elimination (*see* the 1st supplement to the *Théorie des Probababilités*). A longtime before, Mr. Gauss had reduced it to an algorithm. The quantities that he designates by $[bc, 1]$, $[bb, 1], \ldots, [cd, 2]$, etc., are analogues to the Δ of Mr. Cauchy (see *Disquisitio de elementis Palladis*, 1811; or *Theoria Combinationis observationum*, 1828, supplement, page 17). One is even able to encounter an identical march in the eliminations of Legendre (*Nouvelles Recherches sur les Orbites des Comètes*, 1805). This march must have been offered to all the authors, because it is the shortest that one knows for a system of equations of the first degree. It makes part of the education, seeing that it is eminently practical. In fact, when one time the equations are thus restored to contain each at least one unknown, nothing is easier than to write in the first view the value of any one of the unknowns.

One is able to be assured that for m equations among m unknowns, this march requires only $\frac{m-3}{3}(2m^2+5m+6)$ monomial operations, divisions or multiplications, subtractions and additions. For 9 unknowns, for example, 568 operations suffice: a number that one will find very small, if one reflected that the common denominator would be, following the general expression,

$$1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 = 362880 \text{ terms;}$$

and that each of the 9 numerators in containing the same number, would have in all 3628800 terms, each of 10 factors, or 36 million operations.

It would be useless to enter here into the longest developments in this subject. The practitioners will recognize rather, by that which precedes, what advantages one will be able to withdraw or not from these sorts of combinations. The indications given on the reduction of the process of M. Cauchy in the elimination among some equations, sums of products from the given equations by some arbitrary factors, cast one such day on the nature of this process, which one will be able to judge quite better the resources or the faults according to the cases.

In terminating, it is necessary to insist again one time on the difference and even the contradiction which exists between this process and the method of least squares, or each other based on the calculus of probabilities.

P.S. Mr. Cauchy, to which the subject of these remarks had been communicated verbally, appears to have admitted the correctness, because he just proposed to correct, by the method of least squares, the values found by his calculation. The Note that this profound analyst has inserted on this subject into the *Comptes rendu de l'Académie des Sciences*, session of 27 June last,[3] seems however to invite the prompt publication of that which precedes: because the correction of the illustrious author tends nothing less than to double the so painful work of elimination. One has been able to see, in fact, above, that his elimination necessitates exactly the same operations, in a same number, as the method of least squares. To take the approximate values by a process so complex, next to correct them by least squares, returns therefore to make twice all the calculations. Now the resolution of the equations which contain many unknowns, is of all necessity very long, whatever path one may wish to follow; and the practice is objected by all that which increases from it the tedious calculations.

[3]Translator's note: "Mémoire sur l'évaluation d'inconnues déterminées par un grand nombre d'équations approximatives du premier degré. Comptes Rendus Hebd. Séances Acad. Sci. 36, 1114-1122.