About the development of real functions in series using the least squares method

J.P. Gram

Journal für die reine und angewandte Mathematik, v. 94, (1883), pp. 41-73

This is only a partial translation.

Although the thought of using the least squares method for the calculation of the coefficients of an interpolation formula, certainly appears to be almost as old as this method itself, Tchebychef seems to have been the first to carry out this task in all general form. With the solution he gave[1], which is tied to the theory of continued fractions, the result occurs in the form of a series that looked like that of Sturm and Liouville[2] who considered a general class of series somewhat analogous. This analogy, which Bienaymé immediately draws attention to later emerged more clearly through the work of Heine[3]. As well others, such as Plarr[4] and Topler[5], demonstrated that especially the developments according to spherical functions and Fourier series are closely related to the least squares method; however, it seems as though little attention has been paid consistently to this connection. Knowing little about the investigations mentioned above, I have been using for some time the least squares method based on the same kind of series, and I came to the conclusion that this method is the best simplest common point of view for a very large class of developments which are not only the formula of Tchebychef, but also includes the Fourier and similar series. In a work[6], which appeared in the spring of 1879 as a doctoral dissertation, the preliminary results of my investigations in this direction had been laid down. In particular, I not only showed new developments, but also how to use the same method to get a measure of successive approximation. This fact enables one to give a general theory specifically regarding the convergence

[1]Liouville, Journal de Mathématiques, Série II, T. III, 1858, p. 319.

[2]Liouville, Journal de Mathématiques, Série II, T. II p. 220.

[3]Monatsbericht der Berliner Akademie Juni 1866. Siehe auch: Handbuch der Kugelfunctionen 2. Anfl. Bd. 1 Kap. V.

[4]Comptes rendus de lAcad. d. Sc. mai 1857.

[5]Auzeigen der kais. Akad. zu Wien 1876 p. 205.

[6]Om Rakkeudviklinger, bestemte vod Hjailp af de mindste Kvadraters Methode. Kjobenhavn 1879. Host & Son. (122 S. in 8.)

of this kind of series - which I named because of its interpolating character as “interpolation series”. This seems to be all the more of interest to me, because the Dirichlet’s methods used so far have almost exclusively only ever been applied to individual classes of series. - It is the foundations of this theory, which I hereby wish to present to a larger public in a somewhat modified and completed form.

I.

The fundamental problem, the solution of which I will first discuss, is the following.

We are given a series of arguments x and two corresponding quantities o_x, and v_x, all assumed to be real, and v_x positive. In a series with n terms which are known functions of x,

$$y_x = a_1 X_1 + \cdots + a_n X_n, \qquad (1.)$$

the coefficients should then be determined such that the sum $\sum_x v_x(o_x - y_x)^2$ become a minimum. In other words, one should determine the function y such that if o_x are observations and v_x are regarded as associated weights, the sum of the squares of the deviations becomes the smallest. The problem leads to n equations of the form

$$\sum_x v_x X_i o_x = \sum_x v_x X_i y_x, \;\; i = 1, \ldots, n, \qquad (2.)$$

from which one obtains n normal equations for y by inserting the expression (1.). These can be solved using the well-known Gaussian method; but we are taking a different approach.

Let

$$y_x^{(m)} = a_{m,1} X_1 + \cdots + a_{m,m} X_m, \qquad (3.)$$

be the function y, which is obtained if only the first m terms of the series (1.) are used in the adjustment. We also set,

$$s_i = \sum_x v_x X_i o_x, \quad p_{i,k} = p_{k,i} = \sum_x v_x X_i X_k. \qquad (4.)$$

The normal equations for the coefficients of $y^{(m)}$ are

$$\sum_{i=1}^{m} a_{m,i}\, p_{j,i} = s_j, \quad j = 1, \ldots, m. \qquad (5.)$$

Instead of determining the function y directly, we look for the difference

$$y_x^{(m)} - y_x^{(m-1)} = \sum_{i=1}^{m} b_{m,i} X_i,$$

with $b_{m,i} = a_{m,i} - a_{m-1,i}$.

For b, we get $m-1$ equations,

$$\sum_{i=1}^{m-1} b_{m,i}\, p_{j,i} + a_{m,m}\, p_{j,m} = 0, \quad j = 1, \ldots, m-1.$$

From these, one obtains b if $a_{m,m}$ has first been determined from the normal equations (5.). If one denotes by $P^{(m)}$ the symmetric determinant $\sum \pm p_{1,1} \cdots p_{m,m}$ and by $P^{(m)}_{i,k}$ the sub-determinant corresponding to the element $p_{i,k}$ then,

$$a_{m,m} = \frac{\sum_i P^{(m)}_{m,i} s_i}{P^{(m)}},$$

and,

$$\frac{b_{m,1}}{P^{(m)}_{m,1}} = \cdots = \frac{b_{m,m-1}}{P^{(m)}_{m,m-1}} = \frac{a_{m,m}}{P^{(m)}_{m,m}} = \frac{\sum_i P^{(m)}_{m,i} s_i}{P^{(m)} P^{(m-1)}}.$$

If you use the first m conditions with X_1, X_2, …, X_m, in the numerator and denominator, multiplicate and add, it yields,

$$\frac{y^{(m)}_x - y^{(m-1)}_x}{\sum_i P^{(m)}_{m,i} X_i} = \frac{\sum_i P^{(m)}_{m,i} s_i}{P^{(m)} P^{(m-1)}}.$$

or

$$y^{(m)}_x - y^{(m-1)}_x = \frac{\sum_i P^{(m)}_{m,i} s_i}{P^{(m)} P^{(m-1)}} \sum_i P^{(m)}_{m,i} X_i. \qquad (6.)$$

Furthermore, since $y^{(1)} = a_{1,1} X_1 = s_1 / P^{(1)}$, by simple addition for $y^{(n)}$ one gets the following formula,

$$y^{(n)} = \frac{s_1 X_1}{P^{(1)}} + \frac{\sum_i P^{(2)}_{2,i} s_i \sum_i P^{(2)}_{2,i} X_i}{P^{(1)} P^{(2)}} + \cdots + \frac{\sum_i P^{(n)}_{n,i} s_i \sum_i P^{(n)}_{n,i} X_i}{P^{(n-1)} P^{(n)}}. \qquad (7.)$$

The summations should extend to all those values of i for which the determinants retain their meaning.

From this formula, the coefficients a can be determined without difficulty. For our purpose, however, the form obtained here is not the most convenient solution. The function y can also be shown in the form of a series, but not according to the X, but to certain linear functions of the same variables. These functions are so determined that the r^{th} of them also contains only and always the first r of the X, and they thus depend on the order of the members in the series (1.). The general form of these functions, which we will refer below as $\Phi_n(x)$ is the following,

$$\Phi_n(x) = \sum_i P^{(n)}_{n,i} X_i = \begin{vmatrix} p_{1,1} & \cdots & p_{1,n-1} & X_1 \\ \vdots & \ldots & \vdots & \vdots \\ p_{n,1} & \cdots & p_{n,n-1} & X_n \end{vmatrix}. \qquad (8.)$$

so that they appear here in the form of a simply formed determinant. After the introduction of this notation, the series (7.) can be written

$$y^{(n)} = \sum_{i=1}^{n} \frac{A_i}{C_i} \Phi_i(x), \qquad (9.)$$

where

$$A_m = \sum_i P_{m,i}^{(m)} s_i = \sum_x v_x \Phi_m(x) o_x, \quad C_m = P^{(m-1)} P^{(m)}. \qquad (10.)$$

Another expression can be found for the denominator C. By definition

$$\Phi_n(x) = \sum_{i=1}^{n} P_{n,i} X_i,$$

so the sum $\sum_x v_x \Phi_m(x) \Phi_r(x)$ is equal to

$$\sum_{i=1}^{m} P_{m,i} \sum_x v_x X_i \Phi_r(x).$$

But here,

$$\begin{aligned} \sum_x v_x X_i \Phi_r(x) &= P_{r,1} \sum v_x X_i X_1 + \cdots + P_{r,i} \sum v_x X_i X_r, \\ &= \sum_{j=1}^{r} P_{r,j}\, p_{i,j}. \end{aligned}$$

This is 0 if $i < r$ and $P^{(r)}$ if $i = r$, and consequently for $m = r$, you get

$$\sum_x v_x \Phi_r^2(x) = P_{r,r}^{(r)} P^{(r)} = P^{(r-1)} P^{(r)} = C_r, \qquad (11.)$$

and for $m < r$,

$$\sum_x v_x \Phi_m(x) \Phi_r(x) = 0. \qquad (12.)$$

Because of the symmetry, the last equation also applies, of course for $m > r$. Now you can express everything in the series (7.) by means of the functions $\Phi_n(x)$, and the series will take the form[7],

$$y_x^{(n)} = \sum_{i=1}^{n} \frac{\sum v_x \Phi_i(x) o_x}{\sum v_x \Phi_i^2(x)} \Phi_i(x). \qquad (13.)$$

For every adjustment according to a function given in the form (1.) the result can be brought into the typical form (13.). This series can be broken

[7]Tchebychef, Liouville Journal Sér. II, T. III. p. 320.

off at any order and then the result is the adjustment according to a formula which only contains a corresponding number of terms.

Taking into account the given weights, the found function is as close as possible to the observations since the sum $\sum v_x(o_x - y_x)^2$ becomes a minimum. This is obvious since the solution is completely unambiguous and determined, and a maximum if infinite coefficients could occur.

Insofar as the X are available in sufficient numbers, you can continue the series (13.) until you get as many terms as there are observations. In this case, there is complete agreement between the observations and the corresponding values of the function, provided, however, that the nature of the given observations permits this.

In this case, the series (13.) becomes a simple interpolation formula, which can be considered as a transformation of the Lagrangian formula. But if you break off the series at an earlier point, then it gives an approximation formula that includes the observed function. The more terms it contains, the greater the accuracy, and always in such a way that the sum of squares of the deviations becomes the smallest possible. This characteristic property will be shared by all such series using functions $\Phi_n(x)$ that satisfies the condition,

$$\sum v_x \Phi_m(x)\Phi_n(x) = 0, \quad m \neq n. \qquad (14.)$$

You can see this immediately by replacing $\Phi_i(x)$ for X_i in the series (1.). The coefficients will then automatically take the form given in (13.).

To all such developments, which in their most general form are a very large and important type of series, I refer to as "interpolation series", essentially to emphasize their interpolating nature.

For the establishment of interpolation series it is only a matter of determination of suitable systems of "development functions" $\Phi_n(x)$. How these can be specified as determinants for given arguments and weights can be seen from the preceding developments; as can be seen from the studies by Tchebychef and Heine, they can also be shown as approximations of certain continued fractions.

It is not my intention here to examine interpolation series in more detail; I therefore introduce two of them having some practical meaning.

For $x = 0, 1, 2, \ldots, (n-1)$ and $v_x = 1$ you get the series,

$$y = \sum_m (2m+1)\frac{(m!)^2}{(n-m)^{2m+1|1}} \sum_{x=0}^{n-1} \Phi_m(x) o_x \cdot \Phi_m(x), \qquad (15.)$$

where,

$$\Phi_m(x) = (n-1)_m - (m+1)_1(n-2)_{m-1}\frac{x}{1!} + (m+2)_2(n-3)_{m-2}\frac{x^{2|-1}}{2!} - \cdots \qquad (16.)$$

Here the different factors $(a)_p$ denote binomial coefficients.

For $x = 0, 1, 2, \ldots, n$ and $v_x = (n)_x$, we have,

$$y = \sum_m \frac{1}{2^n m! n^{m|-1}} \sum_{x=0}^{n} (n)_x \Phi_m(x) o_x \cdot \Phi_m(x), \qquad (17.)$$

where

$$\Phi_m(x) = n^{m|-1} - (m)_1 (n-1)^{m-1|-1} 2x^{1|-1} + (m)_2 (n-2)^{m-2|-1} 2^2 x^{2|-1} - \cdots \qquad (18.)$$

As other known examples of such series, the trigonometric sine and cosine series can be given by multiples of $\frac{\pi}{n}$, as well as the Lagrangian interpolation formula

$$f(x) = \sum \frac{f(\alpha_i)}{F'(\alpha_i)} \frac{F(x)}{x - \alpha_i},$$

where

$$\Phi_i(x) = \frac{F(x)}{x - \alpha_i}.$$

So far we only have examined the general form of the interpolation series. It is of great interest, however, that the minimum value of $\sum v_x (o_x - y_x^{(n)})^2$ can easily be determined using a simple formula. We will refer to this value below as M_n. One has

$$\sum v_x (o_x - y_x^{(n)})^2 = \sum v_x (o_x - y_x^{(n)}) o_x - \sum v_x (o_x - y_x^{(n)}) y_x^{(n)}.$$

Here the last term is zero because of the general condition

$$\sum v_x (o_x - y_x^{(n)}) X_m = 0,$$

for $m \leq n$. Therefore,

$$M_n = \sum v_x (o_x - y_x^{(n)}) o_x = \sum v_x o_x^2 - \sum v_x y_x^{(n)}) o_x,$$

and here you use the series (13.) for $y^{(n)}$, or with the notation,

$$y^{(n)} = \sum_{i=1}^{n} \frac{A_i}{C_i} \Phi_i,$$

for M_n you get the expression

$$M_n = \sum_x v_x o_x^2 - Q_n, \qquad (19.)$$

where

$$Q_n = \sum_{i=1}^{n} \frac{A_i^2}{C_i}. \qquad (20.)$$

We also get the sum of squares in the form of a series, whose value approaches zero with an increasing number of terms, since all terms of the Q_n series are positive. For each new term that is taken along in the development (13.), a new term is added in the Q series.

The sum M_n must always assume the value zero if the observations and the formally given function y are not such that this is impossible. Insofar as the total weights and the given values of o_x are all of finite sizes, this case can only occur if one or more of the arguments x of $\Phi_n(x)$ (or X_n) disappear. The series gives the value zero always for these arguments. The case $\Phi_n(x) = \infty$ makes the series unusable if $v_x = 0$ or $o_x = 0$ do not occur simultaneously.

II.

So far we have always considered series where the observations are given only for a number of discrete arguments.

We now turn to the consideration of another kind of problem. If one assumes that the arguments x form an arithmetic series whose difference is h, the minimum of $\sum v_x(o_x - y_x)^2$ also becomes the minimum of $\sum v_x(o_x - y_x)^2 h$. But if you ask for this minimum, then the whole development applies, only the factor h is included everywhere under the summations $\sum_x$; otherwise everything remains unchanged. However, if you set h infinitely small in the formulas obtained in this way, then all sums will be transformed into integrals between certain limits by simultaneously accepting v_x as a continuously given positive function, and also o_x as values of a function $f(x)$ given continuously from α to β. One then obtains a series $y^{(n)}$ of such a nature, for which the integral $\int_\alpha^\beta v_x(f(x) - y_x^{(n)})^2 \, dx$ will be minimum.

By repeating the entire previous development, one immediately sees that it remains valid even for this borderline case, and that a single solution is always obtained, which can be represented in the following analogous form (13.),

$$y^{(n)} = \sum_{i=1}^{n} \frac{A_i}{C_i} \Phi_i, \qquad (21.)$$

where Φ satisfies

$$\int_\alpha^\beta v_x \Phi_m(x) \Phi_n(x) \, dx = 0, \quad m \neq n, \qquad (22.)$$

and, by the way, can be represented as determinants. The numerators and denominators of the coefficients have the following form,

$$A_i = \int_\alpha^\beta v_x \Phi_i(x) f(x) \, dx, \quad C_i = \int_\alpha^\beta v_x \Phi_i(x)^2 \, dx. \qquad (23.)$$

For the minimum value of the integral $\int_\alpha^\beta v_x(f(x) - y_x^{(n)})^2 \, dx$, which can be called the square of the mean deviation, there is also a formula analogous to (19.),

$$M_n = \int_\alpha^\beta v_x f^2(x)\, dx - Q_n, \qquad (24.)$$

where Q_n has the same form as before.

[. . .]