JOURNAL

DE

MATHÉMATIQUES

PURES ET APPLIQUÉES

FONDÉ EN 1836 ET PUBLIÉ JUSQU'EN 1874

PAR JOSEPH LIOUVILLE

JULES BIENAYMÉ

Remarques sur les différences qui distinguent l'interpolation de M. Cauchy de la méthode des moindres carrés, et qui assurent la supériorité de cette méthode

Journal de mathématiques pures et appliquées 1re série, tome 18 (1853), p. 299-308.

<http://www.numdam.org/item?id=JMPA_1853_1_18__299_0>

REMARQUES

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

Par M. Jules BIENAYMÉ.

(Extrait des *Comptes rendus de l'Académie des Sciences*, tome XXXVII, séance du 4 juillet 1853.)

« Depuis quelque temps, l'attention de plusieurs observateurs s'est portée sur une *méthode d'interpolation* que M. Cauchy a publiée en 1835 [*], et il semble qu'on ait regardé cette méthode comme ayant quelque chose d'analogue aux avantages de la célèbre méthode des moindres carrés. Il serait fâcheux que les observateurs fussent trompés à cet égard par ce qui a pu être dit des deux méthodes, car elles diffèrent complétement; et si le procédé de M. Cauchy témoigne, comme tout ce qui sort de sa plume, de l'ingénieuse industrie qu'il sait apporter jusque dans les questions pratiques, ce procédé n'en est pas moins tout à fait en contradiction avec les principes du calcul des probabilités. Ce désaccord ne paraît pas être connu, quoiqu'il soit très-facile de l'apercevoir. Mais c'est ce que le temps limité dont les observateurs font le sacrifice à l'analyse, ne leur permet pas de rechercher. Un avertissement peut donc leur être utile; et sans toucher le moins du monde à la valeur que chacun attachera au procédé de M. Cauchy, comme moyen d'interpolation (de séries convergentes surtout), il sera permis de montrer que ce procédé n'est qu'une modification de l'élimination ordinaire entre plusieurs équations du premier degré; modification déjà prescrite par les auteurs qui se sont

[*] *Voir* le tome II de ce Journal, page 193, année 1837.

occupés des moindres carrés, et que M. Gauss a réduite en algorithme; qu'il n'offre aucun degré spécial de probabilité quand on l'applique à des équations plus nombreuses que les inconnues à déterminer; qu'au contraire, il ajoute alors aux risques d'erreur, et n'en assigne pas la mesure; enfin, que, comme moyen d'élimination, il s'appliquerait parfaitement à la méthode des moindres carrés, si par hasard le nombre des inconnues était trop considérable pour qu'on voulût les calculer toutes, et que les dernières ne donnassent d'ailleurs que des termes moindres que les erreurs des quantités observées, supposées seules dans un membre de l'équation. Il est vrai qu'alors rien ne serait plus simple que de supprimer d'avance ces inconnues, que ferait reconnaître l'examen préalable qui doit toujours être fait des équations considérées, avant d'y appliquer un procédé quelconque.

» Pour montrer, aussi simplement que la matière le comporte, comment le procédé de M. Cauchy n'est qu'une modification de l'élimination ordinaire, il suffit de reprendre cette élimination. Soit donc un nombre n d'équations, entre tout autant d'inconnues, de la forme

$$x_1 a_h + x_2 b_h + x_3 c_h + \ldots + x_n l_h = \omega_h. \tag{1}$$

» Si l'on multiplie toutes ces équations par un premier système de facteurs arbitraires $k_{1,h}$, et qu'on ajoute les produits; puis par un second système de facteurs $k_{2,h}$, et qu'on ajoute de même les produits; et qu'on répète cette opération n fois, il est visible qu'on obtiendra n équations de la forme

$$x_1 \mathrm{S}.a_h k_{1,h} + x_2 \mathrm{S}.b_h k_{1,h} + x_3 \mathrm{S}.c_h k_{1,h} + \ldots + x_n \mathrm{S}.l_h k_{1,h} = \mathrm{S}.\omega_h k_{1,h}; \tag{2}$$

$$x_1 \mathrm{S}.a_h k_{i,h} + x_2 \mathrm{S}.b_h k_{i,h} + x_3 \mathrm{S}.c_h k_{i,h} + \ldots + x_n \mathrm{S}.l_h k_{i,h} = \mathrm{S}.\omega_h k_{i,h}. \tag{3}$$

» Quels que soient les facteurs arbitraires k, ces équations pourront remplacer les premières, pourvu que chaque système de facteurs soit différent, afin que les nouvelles équations ne rentrent pas les unes dans les autres. Mais il est palpable que le choix de ces facteurs n'influera pas sur les valeurs finales des inconnues, dont ils disparaîtront entièrement lorsqu'il y aura autant d'inconnues que d'équations primitivement données. Il n'en serait pas de même s'il y avait plus d'équations que d'inconnues; mais c'est un point sur lequel il suffira de revenir plus tard, car rien n'empêche, dans ce cas même, de supposer d'abord autant d'inconnues que l'on voudra, sauf à annuler ensuite les coefficients d'une partie de ces inconnues.

» Maintenant on procédera à l'élimination de x_1 entre la première des nouvelles équations et chacune des $(n-1)$ autres, comme à l'ordinaire, en rendant égaux les coefficients de cette inconnue, et retranchant successivement la première équation de chacune des $(n-1)$ autres.

» Par exemple, pour retrancher de l'équation de rang i, qui a été écrite ci-dessus (3) comme type de toutes, on multiplie la première par le rapport des coefficients de x_1,

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

et l'on obtient sans peine une équation ne renfermant plus x_1 :

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

» Il y aura $(n-1)$ équations de cette forme, et elles ne renfermeront plus que $(n-1)$ inconnues.

» Si l'on fait attention à la composition des coefficients de ces équations, on voit que l'un quelconque

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

peut s'écrire :

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

pourvu qu'on ait dénommé, avec M. Cauchy, par Δc_h, les différences entre parenthèses, et qu'on ait formé toutes ces différences. Ce seront

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

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

$$\cdots\cdots\cdots\cdots$$

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

» Alors les $(n-1)$ équations, entre les $(n-1)$ inconnues, prennent la forme

$$x_2 S.\Delta b_h k_{2,h} + x_3 S.\Delta c_h k_{2,h} + \ldots = S.\Delta \omega_h k_{2,h},$$
$$x_2 S.\Delta b_h k_{i,h} + x_3 S.\Delta c_h k_{i,h} + \ldots = S.\Delta \omega_h k_{i,h}.$$

» Il est manifeste que l'on formerait ces équations, en retranchant la somme (2) des produits par les facteurs $k_{1,h}$ des n équations (1) données, de chacune de ces équations, après avoir multiplié cette somme par $\frac{a_h}{S.a_h k_{1,h}}$. Il viendrait ainsi $(n-1)$ équations de la forme

$$x_2 \Delta b_h + x_3 \Delta c_h + \ldots + x_n \Delta l_h = \Delta \omega_h;$$

et en les multipliant par les facteurs du système $k_{2,h}$, on retomberait sur la première des $(n-1)$ équations déjà obtenues. Les autres dépendraient des systèmes de facteurs désignés par $k_{i,h}$, etc.

» Rien ne vient donc mêler le second système de facteurs avec le premier; et il est placé de même que si l'on ne l'avait introduit qu'après l'élimination de la première inconnue x_1; mais, comme on l'a vu, c'est absolument comme si on l'avait introduit tout d'abord.

» A présent, rien n'est plus aisé que de poursuivre l'élimination des inconnues les unes après les autres. En représentant par Δ^2 des différences formées avec les différences Δ et le système de facteurs $k_{2,h}$, comme les Δ l'ont été avec les coefficients et les facteurs $k_{1,h}$; par exemple,

$$\Delta^2 c_h = \Delta c_h - \Delta b_h \frac{S.\Delta c_h k_{2,h}}{S.\Delta b_h k_{2,h}},$$

$$\Delta^2 \omega_h = \Delta \omega_h - \Delta b_h \frac{S.\Delta \omega_h k_{2,h}}{S.\Delta b_h k_{2,h}},$$

il est clair qu'on parviendra à $(n-2)$ équations entre $(n-2)$ inconnues, de la forme

$$x_3 S.\Delta^2 c_h k_{3,h} + \ldots + S.\Delta^2 l_h k_{3,h} = S.\Delta^2 \omega_h k_{3,h},$$
$$x_3 S.\Delta^2 c_h k_{i,h} + \ldots + S.\Delta^2 l_h k_{i,h} = S.\Delta^2 \omega_h k_{i,h}.$$

» En agissant ensuite de la même manière sur ces équations, on éliminerait encore une inconnue; et il est superflu de pousser plus loin l'opération. Le but est actuellement atteint : c'était de montrer que les facteurs $k_{i,h}$, quels qu'ils soient (M. Cauchy, on le sait, les prend égaux à ± 1), peuvent être introduits dès le commencement de l'opération, sans modifier le moins du monde les résultats. On pouvait penser que M. Cauchy n'introduisant le deuxième système de facteurs qu'après avoir formé les différences Δ, ce système aurait à subir quelque condition spéciale si l'on voulait remonter à la combinaison des équations primitives qui, basée sur ce deuxième système, laisserait pourtant chacun des $k_{2,h}$ tout à fait arbitraire. On pouvait craindre que les $k_{2,h}$ ne vinssent à exiger des facteurs compliqués par les opérations qui conduisent aux équations successives. Il est à présent facile de reconnaître que les choses ne se compliquent pas ainsi, et que l'élimination successive d'une inconnue laisse en dehors des calculs tous les systèmes de facteurs d'un indice plus élevé que l'indice de cette inconnue. Si bien que ces facteurs jouent le même rôle que s'ils venaient d'être introduits arbitrairement.

» Or ils donnent les mêmes résultats que les équations (2) et (3), où ils sont introduits dès l'origine. Et il est très-aisé de voir qu'ils donnent les mêmes résultats, à quelque nombre $m < n$ qu'on réduise ces équations, et les inconnues qu'elles renferment. Car on parviendra successivement (et c'est là le côté ingénieux du procédé, qu'on l'attribue à M. Cauchy ou à M. Gauss [*]), on parviendra aux n équa-

[*] Ce qui distingue surtout le procédé de M. Cauchy, c'est le calcul des restes $\Delta^i \omega_h$, à chaque élimination d'inconnues.

tions :

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

» Que si l'on s'arrête à m inconnues, les premiers termes, dans lesquels entrent ces inconnues, seront précisément les mêmes que si l'on eût pris les équations entières.

» D'un autre côté, on voit très-clairement que ces premiers termes, déduits de n équations de la forme

$$x_1 a_h + x_2 b_h + \ldots + x_n g_h = \omega_h,$$

présentent le résultat ordinaire de l'élimination entre ces équations réduites au nombre m, par la multiplication de m systèmes de n facteurs arbitraires et par l'addition des produits.

» Or on sait, par la méthode des moindres carrés, quels doivent être ces m systèmes chacun de n facteurs, pour que l'erreur finale due aux erreurs partielles des quantités observées ω_h soit un minimum; en d'autres termes, pour que le résultat soit tel, que la somme des carrés des différences entre les ω_h et les premiers membres des n équations données devienne un minimum. Les facteurs $k_{i,h}$, pour satisfaire à cette condition évidemment avantageuse, doivent être les coefficients mêmes des inconnues. Les facteurs de M. Cauchy sont, au contraire, tous égaux à ± 1 : ils ne sauraient donc donner un résultat aussi probable ni aussi avantageux que l'est celui de la méthode des moindres carrés.

» Il y a plus : ces facteurs n'assignent aux résultats aucune probabilité spéciale; car ils prennent leurs signes de manière que toujours $\Delta^{i-1} g_h k_{i,h}$ soit positif, si g_h indique précisément les coefficients de l'inconnue x_i de rang i, avec laquelle apparaissent les facteurs $k_{i,h}$, dans l'ordre d'élimination suivi par M. Cauchy. Or il n'y a là rien qui assigne plutôt une grandeur qu'une autre aux erreurs qu'on laisse subsister.

» Que l'on considère effectivement le premier résultat

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

Les facteurs $k_{1,h}$ sont ± 1, pris de manière que $S.a_h k_{1,h}$ est égale à la somme des valeurs absolues des a_h. Partant

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

est une moyenne entre la plus grande et la plus petite des fractions à dénominateur positif,

$$\frac{\omega_h}{a_h}.$$

» Si ε_h est l'erreur de ω_h, l'erreur de

$$\frac{S\ \omega_h k_{1,h}}{S.a_h k_{1,h}} \text{ sera } \frac{S.\varepsilon_h k_{1,h}}{S\ a_h k_{1,h}},$$

c'est-à-dire une moyenne entre les fractions $\frac{\varepsilon_h}{a_h}$. Il en résulte que l'erreur de

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

sera

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

et comme $\frac{S.\varepsilon_h k_{1,h}}{S.a_h k_{1,h}}$ est tout au plus égale à $\frac{\varepsilon_u}{a_u}$, la plus grande des fractions $\frac{\varepsilon_h}{a_h}$, on aura seulement

$$\varepsilon_h - a_h \frac{\varepsilon_u}{a_u}.$$

Or, à cause du signe et de la grandeur de a_h relativement à la valeur absolue de a_u, il se pourra que $\varepsilon_h - a_h \frac{\varepsilon_u}{a_u}$ fasse une somme supérieure à la plus grande des erreurs ε_h.

» Ce qui vient d'être dit s'applique à tous les degrés de l'opération, de sorte que rien ne garantit que les erreurs n'iront pas en croissant.

» Mais ce n'est pas tout : si l'opération est arrêtée à une inconnue quelconque, elle introduit par sa nature même une autre espèce d'erreur ; puisqu'on néglige alors une suite de termes qui, dans chaque équation, devraient non pas être négligés, mais retranchés des ω_h avant de procéder à l'élimination. Appelant ces quantités négligées δ_h, il est manifeste qu'il arrivera aux δ_h ce que l'on vient de reconnaître pour les ε_h ; et que les combinaisons $\Delta\delta_h$, $\Delta^2\delta_h$, $\Delta^3\delta_h$, etc., pourront grandir et non décroître dans la suite des calculs. Il sera fort difficile d'en être averti ; car les quantités $\Delta\omega_h$, $\Delta^2\omega_h$, etc., que M. Cauchy prend pour indices du terme de l'opération, sont sujettes elles-mêmes à croître et à décroître. On en voit un exemple dans l'interpolation même faite par l'auteur, et publiée dans les *Nouveaux Exercices de Mathématiques*, Prague, 1835. Ainsi l'on n'est pas sûr qu'il faille arrêter les calculs d'après la grandeur de ces indices.

» Il faut se hâter d'ajouter que M. Cauchy n'a proposé sa méthode que pour interpoler des séries dont la convergence est assurée préalablement ; et que dans cette circonstance particulière, les $\Delta\omega_h$, $\Delta^2\omega_h$, etc., iront sans doute en diminuant. Mais son exemple même prouve que ce cas spécial n'est pas exempt de la difficulté signalée ; et cependant la convergence était très-grande. Il est bien clair que cette difficulté affectera à bien plus forte raison l'emploi qu'on pourra faire de sa méthode à des équations de condition, où les inconnues et leurs coefficients ne forment pas une suite très-convergente dans le premier membre. Or on semble aujourd'hui vouloir faire de cette méthode une règle générale, également bonne dans tous les cas.

» On voit que cela n'est pas ; que c'est uniquement un moyen d'élimination qui peut offrir des avantages dans certaines circonstances. L'interpolation est un problème tellement indéterminé, qu'il est bon d'avoir divers procédés, même pour éliminer entre les équations auxquelles on décide qu'il faut s'arrêter. A ce titre, ce sera à l'observateur à discuter le problème qu'il doit résoudre ; et à constater s'il y a pour lui quelque utilité à appliquer le procédé de M. Cauchy, au lieu des méthodes d'interpolation que l'on emploie le plus souvent. Mais quand il voudra obtenir les erreurs minimum, on voit qu'il ne devra pas substituer ce procédé à la méthode des moindres carrés.

» Au surplus, d'après tout ce qui précède, il est visible que les coefficients $k_{i,h}$ peuvent être ceux de la méthode des moindres carrés. Il est donc très-praticable, dans cette méthode, de faire les éliminations successives, en transformant le système d'équations à m inconnues, en un système qui n'aura qu'une équation à m inconnues, une à $(m-1)$ inconnues, une à $(m-2)$, et ainsi de suite, jusqu'à une équation à une seule inconnue, et, si l'on veut, de prendre en considération la grandeur des restes.

» Si donc il se rencontre quelque avantage particulier au procédé, on l'obtiendra sans sacrifier le moins du monde les avantages bien supérieurs de la méthode des moindres carrés. Aussi Laplace avait-il prescrit précisément le même mode d'élimination (*voir* le 1er supplément à la *Théorie des Probabilités*). Longtemps auparavant, M. Gauss l'avait reduit en algorithme. Les quantités qu'il désigne par $[bc, 1]$, $[bb, 1], \ldots, [cd, 2]$, etc., sont analogues aux Δ de M. Cauchy (voir *Disquisitio de elementis Palladis*, 1811; ou *Theoria Combinationis observationum*, 1828, supplément, page 17). On peut même reconnaître une marche identique dans les éliminations de Legendre (*Nouvelles Recherches sur les Orbites des Comètes*, 1805). Cette marche a dû s'offrir à tous les auteurs, parce que c'est la plus courte que l'on connaisse pour un système d'équations du premier degré. Elle faisait partie de l'enseignement, attendu qu'elle est éminemment pratique. En effet, quand une fois les équations sont ainsi ramenées à contenir chacune une inconnue de moins, rien n'est plus facile que d'écrire à la première vue la valeur d'une quelconque des inconnues.

» On peut s'assurer que pour m équations entre m inconnues, cette marche n'exige que $\frac{m-3}{3}(2m^2+5m+6)$ opérations monômes, divisions ou multiplications, soustractions et additions. Pour 9 inconnues, par exemple, il suffit de 568 opérations : nombre qu'on trouvera très-petit, si l'on réfléchit que le dénominateur commun aurait, suivant l'expression générale,

$$1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 = 362\,880 \text{ termes;}$$

et que chacun des 9 numérateurs en contenant le même nombre, il y aurait en tout 3628800 termes, chacun de 10 facteurs, ou 36 millions d'opérations.

» Il serait inutile d'entrer ici dans de plus longs développements à ce sujet. Les praticiens reconnaîtront assez, par ce qui précède, quels avantages on pourra retirer ou non de ces sortes de combinaisons. Les indications données sur la réduction du procédé de M. Cauchy à l'élimination entre des équations, sommes de produits des équations données par des facteurs arbitraires, jettent un tel jour sur la nature de ce procédé, que l'on pourra en juger bien mieux les ressources ou les défauts suivant les cas.

» En terminant, il faut insister encore une fois sur la différence et même la contradiction qui existe entre ce procédé et la méthode des moindres carrés, ou toute autre basée sur le calcul des probabilités.

» *P. S.* M. Cauchy, à qui le sujet de ces remarques avait été communiqué verbalement, paraît en avoir admis la justesse, car il vient de proposer de corriger, par la méthode des moindres carrés, les valeurs trouvées par son calcul. La Note que ce profond analyste a fait insérer à ce sujet dans le *Compte rendu de l'Académie des Sciences*, séance du 27 juin dernier, semble toutefois appeler la prompte publication de ce qui précède : car la correction de l'illustre auteur ne tend rien moins qu'à doubler le travail si pénible de l'élimination. On a pu voir, en effet, ci-dessus, que son élimination nécessite exactement les mêmes opérations, en même nombre, que la méthode des moindres carrés. Prendre des valeurs approchées par un procédé si complexe, puis les corriger par les moindres carrés, revient donc à faire deux fois tous les calculs. Or la résolution des équations qui renferment plusieurs inconnues, est de toute nécessité très-longue, quelque voie que l'on veuille suivre ; et la pratique se refuse à tout ce qui en accroît les fastidieux calculs. »