University of Maryland College Park

Institute for Advanced Computer Studies TR–94–78
Department of Computer Science TR–3307

Gauss, Statistics, and Gaussian Elimination*

G. W. Stewart†

June, 1994
Revised August, 1994

ABSTRACT

This report gives a historical survey of Gauss's work on the solution of linear systems.

*This report is available by anonymous ftp from `thales.cs.umd.edu` in the directory `pub/reports`. It will appear in the proceedings of Interface 94.

†Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742.

Gauss, Statistics, and Gaussian Elimination

G. W. Stewart
Department of Comuter Science and
Institute for Advanced Computer Studies
University of Maryland at College Park

1. Introduction

Everyone knows that Gauss invented Gaussian elimination, and, excepting a quibble, everyone is right.[1] What is less well known is that Gauss introduced the procedure as a mathematical tool to get at the precision of least squares estimates. In fact the computational component in the original description is so little visible, that it takes some doing to see an algorithm in it.

Gaussian elimination, therefore, was not conceived as a general numerical algorithm with applications in statistics and least squares. Rather it was a procedure that sprang from the interface of statistics and computation. Since the full story is known only to the few who have consulted the original sources, I hope my readers will be interested to see how Gauss did things. But there is more than the satisfaction of idle curiosity here. Gauss and Laplace were the premier statisticians of their day, and Gauss alone was the premier numerical analyst. Today we still have something to learn from observing Gauss's practices.

2. Chronicles

The principle of least squares arose from the problem of combining sets of overdetermined equations to form a square system that could be solved for the unknowns. The problem went under the name of the combination of observations, and has been well surveyed by Stigler [23] in his *History of Statistics.* By way of background, I will relate in chronological order the major events in the story of least squares, from Gauss's first discovery to his final treatment in the 1820's.

In his correspondence, Gauss asserted that he had discovered the principle of least squares in 1824 (or 1825, the dates vary). Gauss seems to have had little

[1] The quibble is that in 1759, in the very first paper to appear in his collected works [14], Lagrange gave the basic computational formulas for Gaussian elimination. His purpose, however, was to determine if a critical point was a minimum, not to solve linear equations. There is no indication that the paper had any influence on Gauss, or anyone else.

regard for the principle itself, and even said he thought others must have used it before him. In June of 1828 Gauss [11, v. 10] made the following entry in the little diary of discoveries he kept from 1796 to 1814: "Probability calculus defended against Laplace."[2] Laplace, following Boscovich [1, 16], had suggested that observations be combined by minimizing the sum of the absolute values of the residuals subject to the condition that the residuals sum to zero. Gauss felt that this way of combining observations violated the dictates of probability theory, and his alternative was the first probabilistic justification of least squares.

The following entry in the diary, also dated June 1898, contains the statement: "The problem of elimination resolved in such a way that nothing more can be desired."[3] I take this entry to be the first reference to Gaussian elimination. But a decade was to pass before Gauss published either the probabilistic justification or the elimination procedure.

Although we tend to regard Gauss chiefly as a mathematician, it was as an astronomer that he first made his mark. On New Year's Day of 1801, the astronomer Piazzi discovered the asteroid Ceres. The new planet became unobservable after only nine degrees of an arc had been recorded, and astronomers were faced with problem of determining where to look for it next. Gauss undertook the calculation, using new techniques in physical astronomy and presumably his principle of least squares. At the end of 1801, he predicted where in the heavens the asteroid would be found, and his reputation was made.

Gauss, who was generally slow to publish, began work in 1805 on his *Theoria Motus Corporum Coelestium,* in which he described his techniques for computing orbits and gave his first probabilistic justification of the principle of least squares. He finished in 1806, but his publisher, worried by German losses to Napoleon, insisted he translate the treatise into Latin. In consequence it did not appear until 1809 [2]. In the meantime, Legendre [20] published and named the method of least squares (*la méthode des moindres quarrés*) in an appendix to a memoir appearing in 1805. When the *Theoria Motus* finally appeared, Legendre found that Gauss had claimed the principle for his own, and he took exception. The result was a priority dispute, which need not concern us here. [4]

In the *Theoria Motus*, Gauss had assumed the errors in the observations were normally distributed. In 1811, Laplace [17] used his central limit theorem to give an essentially different justification of least squares. This is not the place to

[2] In the original Latin: *Calculus probabilitatis contra La Place defensus.*

[3] *Problema eliminationis ita solutum, ut nihil amplius desiderari possit.*

[4] Placket [21] gives balanced survey with translations from Gauss's correspondence.

enter into details, but briefly Laplace showed that the solutions of a combination of equations were asymptotically normal and from this concluded that the least squares combination would minimize the mean absolute error in the solutions. Laplace's approach does not readily extend beyond two unknowns.

The final chapter occurred in the 1820's when Gauss [5, 6, 8] published two memoirs on least squares. The first, in two parts, contains yet another justification of least squares — Gauss's famous minimum variance theorem. These papers also contain some nice algorithmics, which will concern us later.

3. The Precision of Estimates

The first appearance of Gaussian elimination in print occurs in Section 182 of the *Theoria Motus.* In order to understand what Gauss is about, we will have to sketch some background.

Gauss (after a linearization) considers the model[5]

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e},$$

where $\mathbf{X}$ is $n \times p$. The errors e_i are assumed to be independent random variables with common distribution $\varphi(e)$. Gauss introduces the function

$$\varphi(y_1 - \mathbf{x}_1^{\mathrm{T}}\mathbf{b})\varphi(y_2 - \mathbf{x}_2^{\mathrm{T}}\mathbf{b}) \cdots \varphi(y_n - \mathbf{x}_n^{\mathrm{T}}\mathbf{b}), \tag{3.1}$$

where the $\mathbf{x}_i^{\mathrm{T}}$ are the rows of $\mathbf{X}$ and uses a Bayesian argument with a uniform prior to argue that the value of $\mathbf{b}$ that maximizes (3.1) is the most probable value of the unknowns.

Gauss now supposes the distribution of the e_i is normal; that is, $\varphi(e) \propto e^{-h^2 e^2}$. He identifies the parameter h with the precision[6] of $\mathbf{y}$. The function (3.1) now becomes proportional to

$$e^{-h^2 \Omega}, \tag{3.2}$$

where

$$\Omega = (\mathbf{y} - \mathbf{X}\mathbf{b})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\mathbf{b})$$

[5] We will make free use of matrices in what follows, but only as a means of abbreviating Gauss's scalar equations.

[6] We must not use terms like variance or standard deviation here. The number h is simply a parameter in a specific distribution. Only in the *Theoria Combinationis* will Gauss introduce the second moment of a general distribution as a measure of variation.

is the residual sum of squares. Thus, Gauss's most probable value is obtained by minimizing the residual sum of squares, which justifies the principle of least squares. The normal equations can be derived as usual by differentiation.

Gauss next turns to the problem of estimating the precision of the least squares estimates. His technique is to integrate all but the last unknown out of (3.2), after which the precision can be read off. However, to perform the required integrations Ω must be expressed in a special form, and the tool for arriving at that form is Gaussian elimination.

The procedure as given by Gauss is the following. Let

$$u_1 = \frac{1}{2}\frac{\partial\Omega}{\partial b_1} \equiv r_{11}b_1 + r_{12}b_2 + \cdots + r_{1p}b_p - s_1, \tag{3.3}$$

and let

$$\Omega_1 = \Omega - \frac{u_1^2}{r_{11}}. \tag{3.4}$$

Then clearly the derivative of Ω_1 with respect to b_1 is zero, so that Ω_1 is independent of b_1.

One more step will illustrate the general procedure. Set

$$u_2 = \frac{1}{2}\frac{\partial\Omega_1}{\partial b_2} \equiv r_{22}b_2 + r_{23}b_3 + \cdots + r_{2p}b_p - s_2.$$

Then

$$\Omega_2 = \Omega_1 - \frac{u_2^2}{r_{22}}$$

is independent of b_1 and b_2. Continuing in this manner we arrive at the decomposition

$$\Omega = \frac{u_1^2}{r_{11}} + \frac{u_2^2}{r_{22}} + \cdots + \frac{u_p^2}{r_{pp}} + \rho,$$

in which u_i is independent of $b_1, \ldots, b_{i-1}$ and ρ is constant.

Gauss now considers the expression

$$e^{-h^2\Omega} \propto \exp\Big(-h^2\frac{u_1^2}{r_{11}}\Big) \cdot \exp\Big(-h^2\frac{u_2^2}{r_{22}}\Big) \cdots \exp\Big(-h^2\frac{u_p^2}{r_{pp}}\Big).$$

and integrates with respect to b_1 over the real line. Since the last $p-1$ factors in this expression are free of b_1, they remain unchanged by the integration. The first factor integrates to a constant. Thus Gauss is left with a distribution proportional to

$$e^{-h^2\Omega_1} \propto \exp\Bigl(-h^2\frac{u_2^2}{r_{22}}\Bigr)\cdots\exp\Bigl(-h^2\frac{u_p^2}{r_{pp}}\Bigr),$$

which is free of b_1. Continuing this process of integrating out the parameters b_i, Gauss finds that the distribution of b_p is proportional to

$$\exp\Bigl(-h^2\frac{u_p^2}{r_{pp}}\Bigr),$$

where

$$u_p = r_{pp}b_p - s_p.$$

Gauss concludes that the most probable value of b_p, obtained by setting $u_p = 0$, is

$$\hat{b}_p = \frac{s_p}{r_{pp}}$$

and its precision is

$$\frac{h}{\sqrt{r_{pp}}}.$$

Gauss now goes on to show that if you write the normal equations in the form

$$\mathbf{A}\mathbf{b} = \mathbf{c} \tag{3.5}$$

and express $\mathbf{b}$ as a function of $\mathbf{c}$ in the form

$$\mathbf{b} = \mathbf{V}\mathbf{c}, \tag{3.6}$$

then the (p,p)-element of $\mathbf{V}$ is $\frac{1}{r_{pp}}$. Since the resulting expression for the precision clearly does not depend on the position of the unknown, Gauss concludes that the precision of *any* of the estimates $\hat{b}_i$ is $h\sqrt{v_{ii}}$.

It is ironic that the *Theoria Motus* should have become the principle reference for Gaussian elimination as a computational tool. As we have seen, Gauss used elimination to give a derivation of one of the most important results of linear regression theory. He was certainly aware of the computational consequences of his elimination procedure, and promised to describe them in a later work. But computational considerations are absent from the *Theoria Motus* itself. Gauss merely points out that the normal equations can be solved by ordinary elimination (*eliminatio vulgaris*), presumably a variant of what we now call Gauss–Jordan elimination. An extension, which Gauss will later call general elimination (*eliminatio indefinite*), can be used to pass from the normal equations (3.5) to the inverse system (3.6).

4. The Scalar Connection

In 1810, in *Disquisitio de Elementis Ellipticis Palladis* [3], Gauss gave the numerical details of his algorithm and illustrated it with an example. The usual scalar formulas for Gaussian elimination can be derived by translating the original algorithm into operations on the matrix of second derivatives of Ω. Specifically, if we set

$$a_{ij} = \frac{1}{2}\frac{\partial^2\Omega}{\partial b_i \partial b_j},$$

then we see from (3.3) that $a_{11} = r_{11}$ and from (3.4) that

$$\Omega_1 = \Omega - \frac{1}{a_{11}}\Big(\frac{1}{2}\frac{\partial\Omega}{\partial b_1}\Big)^2.$$

It follows that

$$a_{ij}^{(1)} \equiv \frac{1}{2}\frac{\partial^2\Omega_1}{\partial b_i \partial b_j} = a_{ij} - \frac{a_{i1}a_{1j}}{a_{11}}.$$

In the expression on the right, we recognize the formulas for performing one step of Gaussian elimination, as we understand it today, on a matrix whose elements are a_{ij}. This is essentially the algorithm Gauss describes in the *Disquisitio.*

To complete the solution of the normal equations by Gaussian elimination, note that since

$$\Omega = \frac{u_1^2}{r_{11}} + \frac{u_2^2}{r_{22}} + \cdots + \frac{u_p^2}{r_{pp}} + \rho,$$

the function Ω assumes its minimum value ρ when

$$u_1 = u_2 = \cdots = u_p = 0.$$

Since

$$0 = u_p = r_{pp}b_p - s_p$$

is a linear equation involving only b_p, it can be solved immediately for b_p. Having determined b_p, one can solve for b_{p-1} from the equation

$$0 = u_{p-1} = r_{p-1,p-1}b_{p-1} + r_{p-1,p}b_p - s_{p-1}.$$

Continuing in this manner, we can determine estimates for all the unknowns b_i. This of course is nothing more than the back substitution phase of Gaussian elimination.

5. The Matrix Connection

The above description of the algorithm is incomplete, in the sense that it does not give formulas for the constant parts s_i of the functions u_i. To see where they come from, it will be useful to express the algorithm in terms of matrices.

The function Ω can be written in the form

$$\begin{aligned}\Omega &= (\mathbf{b}^{\mathrm{T}} \;\; -1)\begin{pmatrix}\mathbf{X}^{\mathrm{T}}\mathbf{X} & \mathbf{X}^{\mathrm{T}}\mathbf{y}\\ \mathbf{y}^{\mathrm{T}}\mathbf{X} & \mathbf{y}^{\mathrm{T}}\mathbf{y}\end{pmatrix}\begin{pmatrix}\mathbf{b}\\ -1\end{pmatrix}\\ &\equiv (\mathbf{b}^{\mathrm{T}} \;\; -1)\begin{pmatrix}\mathbf{A} & \mathbf{c}\\ \mathbf{c}^{\mathrm{T}} & \eta\end{pmatrix}\begin{pmatrix}\mathbf{b}\\ -1\end{pmatrix}\end{aligned}$$

If we set

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1p}\\ 0 & r_{22} & \cdots & r_{2p}\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & r_{pp}\end{pmatrix} \quad \text{and} \quad \mathbf{s} = \begin{pmatrix} s_1\\ s_2\\ \vdots\\ s_p\end{pmatrix},$$

where the r's and s's are from the definitions of the functions u [see (3.3)], then it is easy to verify that

$$\begin{pmatrix}\mathbf{A} & \mathbf{c}\\ \mathbf{c}^{\mathrm{T}} & \eta\end{pmatrix} = \begin{pmatrix}\mathbf{R}^{\mathrm{T}} & 0\\ \mathbf{s}^{\mathrm{T}} & \rho\end{pmatrix}\begin{pmatrix}\mathbf{D}^{-1} & 0\\ 0 & \rho^{-1}\end{pmatrix}\begin{pmatrix}\mathbf{R} & \mathbf{s}\\ 0 & \rho\end{pmatrix},$$

where

$$\mathbf{D} = \mathrm{diag}(r_{11}, r_{22}, \ldots, r_{pp}).$$

Thus Gaussian elimination, as practiced by Gauss, amounts to factoring the *augmented cross-product matrix* into a lower triangular matrix, a diagonal matrix, and the transpose of the lower triangular matrix. It is common practice today to work with the augmented cross-product matrix.

The vector $\mathbf{u}$ whose components are the functions u_i can be written in the form

$$\mathbf{u} = \mathbf{R}\mathbf{b} - \mathbf{s}.$$

The process sketched above of setting the u_i to zero and back-solving amounts to solving the triangular system

$$\mathbf{R}\mathbf{b} = \mathbf{s}.$$

6. The Computation of Variances

Writing in 1821, Gauss [4] summarized his and Laplace's justifications of least squares as follows.

> From the foregoing we see that the two justifications each leave something to be desired. The first depends entirely on the hypothetical form of the probability of the error; as soon as that form is rejected, the values of the unknowns produced by the method of least squares are no more the most probable values than is the arithmetic mean in the simplest case mentioned above. The second justification leaves us entirely in the dark about what to do when the number of observations is not large. In this case the method of least squares no longer has the status of a law ordained by the probability calculus and has only the simplicity of the operations it entails to recommend it.

In the *Pars Prior* of his memoir *Theoria Combinationis Observationum Erroribus Minimis Obnoxiae* [7], Gauss resolved the dilemma by introducing the notion of mean square error as a measure of variance and showing that among all linear combinations of the observations that produced exact estimates in the absence of error the least squares estimates have least mean square error.

In the *Pars Posterior* of the *Theoria Combinationis* [6], Gauss addresses the problem of computing variances. He points out that his elimination method gives only the variance of the last unknown. Since (he continues) a general elimination to invert the normal equations is expensive, some calculators have adopted the practice of performing the elimination with another unknown placed last.[7] Gauss says that he will give a better way.

Gauss actually gives two solutions to the problem. In the first he shows that if one inverts the system $\mathbf{Rb} = \mathbf{s}$ to get $\mathbf{Ts} = \mathbf{b}$, then the matrix $\mathbf{V}$ obtained by passing from (3.5) to (3.6) can be written

$$\mathbf{V} = \mathbf{TDT}^{\mathrm{T}}.$$

Thus the diagonal elements of $\mathbf{V}$ can be computed as a weighted sum of squares of the rows of $\mathbf{T}$. Gauss gives two algorithms for computing $\mathbf{T}$, one of them particularly advantageous when only a few variances are to be computed.

[7] Laplace, for example, recommended a similar procedure in the first supplement to his *Théorie Analytique des Probabilités* [18].

The second method is a very general result for computing the variance of an arbitrary linear combination

$$t = \mathbf{g}^{\mathrm{T}}\mathbf{b} + \kappa$$

of the unknowns $\mathbf{b}$. Specifically, if we pass from the variables $\mathbf{b}$ to the variables $\mathbf{u}$, so that t assumes the form

$$t = \mathbf{h}^{\mathrm{T}}\mathbf{u} + \hat{t},$$

then $\hat{t}$ is the value of t at the least squares estimates of the unknowns,[8] and its variance is proportional to

$$\mathbf{h}^{\mathrm{T}}\mathbf{D}\mathbf{h}.$$

Moreover, $\mathbf{h}$ may be obtained by solving the triangular system

$$\mathbf{R}^{\mathrm{T}}\mathbf{h} = \mathbf{g}.$$

Thus Gauss reduces the problem of computing a variance to that of solving a triangular system.

A modern practice in numerical linear algebra is to compute a matrix decomposition and then use it in a variety of computations. Although it would be anachronistic to call Gauss a decompositionalist, he calculated like one. The results of his elimination serve as a computational platform from which both estimates and variances can be obtained.

7. Computational Complexity

Did Gaussian elimination represent an improvement over the practices of the day? If we assume that people were using Gauss–Jordan elimination to solve systems, they would have performed roughly $\frac{1}{2}p^3$ multiplications and about the same number of additions. Gaussian elimination, on the other hand, requires about $\frac{1}{6}p^3$ multiplications and additions. Thus Gaussian elimination represents an improvement of a factor of about three.

If variances are required, the inversion of the normal equations by Gauss-Jordan elimination would cost an additional $\frac{1}{3}p^3$ multiplications and additions for a total of $\frac{5}{6}p^3$. With Gauss's approach the total is $\frac{1}{3}p^3$, an improvement by a factor $\frac{5}{2}$.

[8] It has been asserted [22] that Gauss established that $\hat{t}$ enjoyed the same minimum variance properties as the components of $\hat{\mathbf{b}}$. Although the result is true, Gauss never proved it.

In an age in which a workstation can solve a system of order 100 with barely a hiccup, it is easy to be cavalier about factors of three. To see what it might have meant to people who had to do their calculations by hand, consider the following quote from *A Treatise on the Adjustment of Observations* published in 1884 by T. W. Wright [24, p. 173]:

> Dr. Hügel, of Hessen, Germany, states that he has solved 10 normal equations in from 10–12 hours, using a log. table, but that 29 equations took him seven weeks.

Without Gaussian elimination Dr. Hügel's twelve hours would have stretched to a day and a half, and his seven weeks to almost half a year.

8. Notation

Gauss, like most mathematicians of his time, made sparing use of subscripts and superscripts, prefering to use primes or sequences of letters to distinguish variables. For example, Gauss writes his linear model in the form

$$\begin{aligned} v &= ax + by + cx + \text{etc.} + l \\ v' &= a'x + b'y + c'x + \text{etc.} + l' \\ v'' &= a''x + b''y + c''x + \text{etc.} + l'' \text{ etc.} \end{aligned}$$

Here x, y, z, etc. are the unknowns we have been denoting by b_i and the v's are the errors. Although this expansive notation appears awkward to us, in Gauss's hands it could be quite expressive. For example, here (slightly edited) is how he writes the normal equations.

$$\begin{aligned} 0 &= [aa]x + [ab]y + [ac]z + \text{etc.} + [al] \\ 0 &= [ab]x + [bb]y + [bc]z + \text{etc.} + [bl] \\ 0 &= [ac]x + [bc]y + [cc]z + \text{etc.} + [cl] \text{ etc.} \end{aligned}$$

Note the elegant way in which the notation $[ab]$ suggests a sum of products from the a and b columns.

Gauss's notation for elimination is equally well considered. The following is from the *Supplementum* [8] to the *Theoria Combinationis*

$$
\begin{aligned}
&[bb,1] = [bb] - \tfrac{[ab]^2}{[aa]} \\
&[bc,1] = [bc] - \tfrac{[ab][ac]}{[aa]} \\
&[bd,1] = [bd] - \tfrac{[ab][ad]}{[aa]} \\
&\text{etc.} \\
&[cc,2] = [cc] - \tfrac{[ac]^2}{[aa]} - \tfrac{[bc,1]^2}{[bb,1]} \\
&[cd,2] = [cd] - \tfrac{[ac][ad]}{[aa]} - \tfrac{[bc,1][bd,1]}{[bb,1]} \\
&\text{etc.} \\
&[dd,3] = [dd] - \tfrac{[ad]^2}{[aa]} - \tfrac{[bd,1]^2}{[bb,1]} - \tfrac{[cd,2]^2}{[cc,1]}
\end{aligned}
$$

Here as above, a pair of letters indicates the position in the normal equations. The appended numerals indicate the level of elimination. Incidentally, this seems to be the first appearance of the inner product form of the algorithm, in which the matrix $\mathbf{R}$ is generated row by row. It is the preferred form for hand calculation, since one need only record an array of $\frac{1}{2}p^2$ numbers.

9. Legacy

The casting of Gauss's results in matrix notation in some sense trivializes them. With our knowledge of matrix algebra, we can leap ahead to results that researchers of Gauss's time could only arrive at by more pedestrian routes. Yet we must be careful not to be patronizing. Gauss and his successors accomplished a great deal with their techniques and notation.

For example, Gauss's presentation of his algorithm as elimination in a quadratic form strikes us as unusual today. Yet it was the first of many reductions of quadratic and bilinear forms that later became our familiar matrix decompositions, including among others the LU decomposition, the Jordan canonical form, and the singular value decomposition. As Kline points out in his book *Mathematical Thought from Ancient to Modern Times* [13, Ch. 33], by the time the use of matrices had become widespread, many of the principal results of matrix theory had already been established.

Gauss's algorithms, written in his notation, survived into the twentieth century, especially in books on geodesy. Thereafter, as people began to use present-day notation, his contributions became less visible. By 1959, when I first began working with computers, Gaussian elimination had come to mean any triangularization of a system of equations, symmetric or nonsymmetric, followed by a back substitution, and none of us had an idea of what Gauss had actually done.

Yet what he did is worth recalling. Gauss worked with real-life problems and got his hands dirty solving them. He always looked for the best, most efficient algorithm; and when he had it, he expressed it in a clean notation that suggested how to use it. These virtues are no less important today than in Gauss's time.

References

[1] R. J. Boscovich and C. Maire. *De Litteraria Expeditione per Pontificiam ditionem ad dimetiendas duas Meridiani graduss.* Palladis, Rome, 1755. Cited in [23].

[2] C. F. Gauss. *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium.* Perthes and Besser, Hamburg, 1809. Cited and reprinted in [11, v. 7, pp. 1–261]. English translation by C. H. Davis [10]. French and German translations of Book II, Part 3 in [9, 12].

[3] C. F. Gauss. Disquisitio de elementis ellipticis Palladis. *Commentatines societatis regiae scientarium Gottingensis recentiores*, 1, 1810. Cited and reprinted in [11, v. 6, pp. 1–64]. French translation of §§13–14 in [9]. German translation of §§10–15 in [12].

[4] C. F. Gauss. Anzeige: Theoria combinationis observationum erroribus minimis obnoxiae: Pars prior. *Göttingische gelehrte Anzeigen*, 33:321–327, 1821. Cited and reprinted in [11, v. 4, pp. 95–100].

[5] C. F. Gauss. Anzeige: Theoria combinationis observationum erroribus minimis obnoxiae: Pars posterior. *Göttingische gelehrte Anzeigen*, 32:313–318, 1823. Cited and reprinted in [11, v. 4, pp. 100–104].

[6] C. F. Gauss. Theoria combinationis observationum erroribus minimis obnoxiae: Pars posterior. *Commentatines societatis regiae scientarium Gottingensis recentiores*, 5, 1823. Cited and reprinted in [11, v. 4, pp. 27–53]. French and German translations in [9, 12].

[7] C. F. Gauss. Theoria combinationis observationum erroribus minimis obnoxiae: Pars prior. *Commentatines societatis regiae scientarium Gottingensis*

recentiores, 5, 1823. Cited and reprinted in [11, v. 4, pp. 1–26]. French and German translations in [9, 12].

[8] C. F. Gauss. Supplementum theoriae combinationis observationum erroribus minimis obnoxiae. *Commentatines societatis regiae scientarium Gottingensis recentiores*, 6, 1828. Cited and reprinted in [11, v. 4, pp. 55–93]. French and German translations in [9, 12].

[9] C. F. Gauss. *Méthode des Moindres Carres*. Ballet–Bachelier, Paris, 1855. Translation by J. Bertrand of various works of Gauss on Least Squares.

[10] C. F. Gauss. *Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections*. Little, Brown, and Company, 1857. Translation by Charles Henry Davis of *Theoria Motus* [2]. Reprinted by Dover, New York, 1963.

[11] C. F. Gauss. *Werke*. Kñiglichen Gesellschaft der Wissenschaften zu Göttingen., 1870–1928.

[12] C. F. Gauss. *Abhandlungen zur Methode der kleinsten Quadrate*. P. Stankeiwica', Berlin, 1887. Translation by A. Borsch and P. Simon of various works of Gauss on Least Squares.

[13] M. Kline. *Mathematical Thought from Ancient to Modern Times*. Oxford University Press, New York, 1972.

[14] J.-L. Lagrange. Researches sur la métode de maximis et minimis. *Miscellanea Taurinensi*, 1, 1759. Cited and reprinted in [15, v. 1, pp. 1–16].

[15] J.-L. Lagrange. *Œvres de Langrange*. Gauthier–Villars, Paris, 1867–1892.

[16] P. S. Laplace. Sur quelques points du systéme du monde. *Memoires de l'Academie des Sciences de Paris*, 1789. Cited and reprinted in [19, v. 11, pp. 475–558].

[17] P. S. Laplace. Mémoire sur les intégrales définies et sur application aux probabilités. *Memoires de l'Academie des Sciences de Paris*, 11, 1810–1811. Cited and reprinted in [19, v. 12, pp. 355–412].

[18] P. S. Laplace. *Théorie Analytique des Probabilités*. Courcier, Paris, third edition, 1820. Reprinted in [19, v. 7].

[19] P. S. Laplace. *Œvres Compeétes*. Gauthier–Villars, Paris, 1878–1912.

[20] A. M. Legendre. *Nouvelle méthodes pour la détermination des orbites des comètes*. Courcier, Paris, 1805. Cited in [23], where the appendix on least squares is reproduced.

[21] R. L. Plackett. The discovery of the method of least squares. *Biometrika*, 59:239–251, 1972.

[22] H. L. Seal. The historical development of the Gauss linear model. *Biometrika*, 54:1–24, 1967.

[23] S. M. Stigler. *The History of Statistics*. Harvard University Press, Cambridge, Massachusetts, 1986.

[24] T. W. Wright. *A Treatise on the Adjustment of Observations*. Van Nostrand, New York, 1884.