and therefore the corresponding equation (14) is actually

$$\left(\bar{X}^{-1} + \frac{1}{m}\delta\right)^2 = \bar{X}^{-2} - \frac{1}{m}\Delta + \frac{1}{m}0[(\bar{X}^{-1}\delta)^2] + \frac{1}{m^2}\delta^2.$$

Now, since the third term of the right hand side of this equation is to be neglected, one concludes that the removal of the last term does not cause any additional inaccuracy, if m is large enough. It is sufficient to choose m of the same order of magnitude as the largest eigenvalue of $\bar{X}$ squared or the reciprocal of the smallest eigenvalue of A. Compute the matrix $\left(\bar{X}^{-2} - \frac{1}{m}\Delta\right)^{\frac{1}{2}}$ by means of the algorithm (15); subtract $\bar{X}^{-1}$ from this result; and finally, multiply by m in order to obtain an approximation for the correction (11). Of course, this procedure may be repeated and thus set up a quadratically convergent algorithm which, moreover, is self-correcting.

Summary. It has been proved that for any real $(n \times n)$-matrix A with only positive eigenvalues the algorithm (1), with an initial matrix $X^{(0)} = kI$, converges quadratically to the matrix $A^{-\frac{1}{2}}$ with positive eigenvalues.

In a numerical case this algorithm, if continued indefinitely, may be divergent due to round-off errors, whose influence may increase in geometrical progression. This makes it necessary to stop the process as soon as the difference between two successive results no longer decreases; of course, it is also desirable to have some additional accuracy in the numerical computation to take care of the round-off errors.

Any approximation to $A^{-\frac{1}{2}}$ sufficiently accurate can be improved successively, the rate of convergence of the procedure being quadratic. Each step, however, involves either the solution of a system of n^2 linear equations or the extraction of the square-root of a matrix, which may be achieved by a quadratically convergent iteration process.

Kasarmikatu 2B11
Helsinki, Finland

On Gauss' Speeding Up Device in the Theory of Single Step Iteration

By Alexander M. Ostrowski

1. In this paper we consider throughout *real* numbers, vectors and matrices. In order to solve the linear system

$$(1) \qquad \sum_{\nu=1}^{n} a_{\mu\nu}x_\nu = y_\mu, \quad a_{\mu\mu} = A_\mu \quad (\mu = 1, \cdots, n)$$

Received in the present version 7 October 1957. This work was performed under a contract of the National Bureau of Standards with the American University and University of California at Los Angeles.

with the matrix A, $A_\mu \neq 0$ $(\mu = 1, \cdots, n)$ and non-vanishing determinant, by the *single step iteration* we form, starting from an arbitrary (row) vector ξ_0, a sequence of vectors

$$\xi_k = (x_1^{(k)}, \cdots, x_n^{(k)}) \quad (k = 1, 2, \cdots) \tag{2}$$

obtained in the following way: For any integer $k (k \geqslant 0)$ choose a value N_k of the "leading index" from the indices $1, \cdots, n$; then if

$$\rho_k = (r_1^{(k)}, \cdots, r_n^{(k)}) \quad (k = 0, 1, 2, \cdots) \tag{3}$$

is the k-th *residual vector* defined by

$$r_\mu^{(k)} = \sum_{\nu=1}^{n} a_{\mu\nu} x_\nu^{(k)} - y_\mu \quad (\mu = 1, \cdots, n;\ k = 0, 1, \cdots), \tag{4}$$

we put

$$x_\mu^{(k+1)} = x_\mu^{(k)} \quad (\mu \neq N_k), \qquad x_{N_k}^{(k+1)} = x_{N_k}^{(k)} - \frac{r_{N_k}^{(k)}}{A_{N_k}}. \tag{5}$$

In studying this iteration we can and will restrict ourselves to the case where all y_μ vanish, since we can always by a convenient change of the origin make all $y_\mu = 0$, without changing the ρ_k.

2. We consider in what follows only the case where the matrix of the system (1) is *symmetric* and the quadratic form

$$K(\xi) = \sum_{\mu,\nu=1}^{n} a_{\mu\nu} x_\mu x_\nu = \xi A \xi' \tag{6}$$

defined for an arbitrary vector $\xi = (x_1, \cdots, x_n)$, is *positive definite*. In this case it is well known and immediately verified that if the vector ξ_{k+1} is obtained from the vector ξ_k by the transformation (5), we have

$$K(\xi_{k+1}) = K(\xi_k) - (r_{N_k}^{(k)})^2 / A_{N_k}. \tag{7}$$

In using (7) it was proved by Seidel, 1874 [13], that the single step iteration is always convergent if N_k is chosen at each step so that

$$(r_{N_k}^{(k)})^2 / A_{N_k} = \max_\mu (r_\mu^{(k)})^2 / A_\mu \tag{8}$$

This is Seidel's relaxation procedure.

This special rule goes back to F. R. Helmert, 1872 [7]. The relaxation rule indicated previously by Gauss [4] and Gerling [6] is different, as is the one proposed by Southwell [14], but the rule (8) is apparently the most advantageous one. Cf. the discussion in [9], p. 158–9.

On the other hand Schmeidler [12] and Reich [11] proved in using (7) that the single step procedure is convergent in the *cyclic* case when N_k runs periodically through all indices $1, \cdots, n$.

3. Gauss [4], [1], [6] and [15] proposed the following modification of the

above procedure in order to speed up the convergence. Put

$$(9)\qquad \begin{aligned} x_\nu &= z_\nu - z_0, \quad a_{0\nu} = a_{\nu 0} = -\sum_{\mu=1}^{n} a_{\mu\nu} \quad (\nu = 1, \cdots, n) \\ a_{00} &= A_0 = -\sum_{\nu=1}^{n} a_{0\nu} = \sum_{\mu,\nu=1}^{n} a_{\mu\nu}, \end{aligned}$$

where z_0 can be arbitrarily chosen. Then the system (1) can be written in the form (assuming $y_\mu = 0$)

$$(10)\qquad \sum_{\nu=0}^{n} a_{\mu\nu} z_\nu = 0 \quad (\mu = 0, 1, \cdots, n),$$

where the first equation is, of course, not independent of the last n equations but is useful for the sake of uniformity and for checking purposes.

In particular A_0 is positive since by (9) A_0 is the value of the quadratic form (6) for $x_\nu = 1$ $(\nu = 1, \cdots, n)$.

4. From a solution $(z_0, z_1, \cdots, z_n)$ of the system (10) we obtain at once by (9) the solution $(x_1, \cdots, x_n)$ of the system (1). The idea of Gauss is now to apply the procedure described in (4) and (5) to the system (10). If we obtain then, starting from a vector $\zeta_0 = (z_0^{(0)}, z_1^{(0)}, \cdots, z_n^{(0)})$, a sequence of vectors

$$\zeta_k = (z_0^{(k)}, z_1^{(k)}, \cdots, z_n^{(k)}),$$

we consider at the same time the corresponding vectors

$$\xi_k = (z_1^{(k)} - z_0^{(k)}, \cdots, z_n^{(k)} - z_0^{(k)}).$$

If then in the passage from ζ_k to ζ_{k+1} the leading index $N_k \neq 0$, we have

$$(11)\qquad \sum_{\nu=0}^{n} a_{\mu\nu} z_\nu^{(k)} = \sum_{\nu=1}^{n} a_{\mu\nu}(z_\nu^{(k)} - z_0^{(k)}) = r_\mu^{(k)} \quad (\mu = 1, \cdots, n),$$

$$\begin{aligned} z_{N_k}^{(k+1)} &= z_{N_k}^{(k)} - \frac{\sum_{\nu=0}^{n} a_{N_k\nu} z_\nu^{(k)}}{A_{N_k}} = z_{N_k}^{(k)} - \frac{r_{N_k}^{(k)}}{A_{N_k}}, \\ z_\mu^{(k+1)} &= z_\mu^{(k)} \quad (\mu \neq N_k). \end{aligned}$$

Since here $z_0^{(k+1)} = z_0^{(k)}$, we see that the corresponding n-dimensional vectors ξ_k, ξ_{k+1} are connected exactly by the formulae (5), so that in this case there is no essential change compared with the original method.

5. If, however, $N_k = 0$, then only $z_0^{(k)}$ is changed and therefore *all components* $x_1, \cdots, x_n$ are changed by the same amount. In this case we have obviously a new possibility and the question arises whether in this case the convergence is indeed speeded up. Of course, by 'convergence' in this case is not meant the convergence of the vectors ζ_k but that of the corresponding vectors ξ_k. This question is ap-

parently not as yet settled, as widely contradictory opinions are to be found in the literature; see [14] and [3].

Our conclusions are stated at the start of section 8 and at the end of the paper.

6. In what follows we will say that the two $(n+1)$-dimensional vectors $\xi = (z_0, z_1, \cdots, z_n)$ and $\zeta' = (z_0', z_1', \cdots, z_n')$ are equivalent, if we have $z_\nu - z_0 = z_\nu' - z_0'$. In the class of vectors equivalent to ζ there exists a *reduced* one: $\hat{\zeta} = (0, x_1, \cdots, x_n)$, and the corresponding n-dimensional vector $\xi = (x_1, \cdots, x_n)$ is uniquely determined.

If we define the component of the residual vector corresponding to the index 0 by

$$r_0^{(k)} = \sum_{\nu=0}^{n} a_{0\nu} z_\nu^{(k)} = \sum_{\nu=1}^{n} a_{0\nu}(z_\nu^{(k)} - z_0^{(k)}) = -\sum_{\nu=1}^{n} r_\nu^{(k)}, \tag{12}$$

we see from (11) and (12) that the residual vector for the system (10) does not depend on the component z_0 but only on the corresponding vector ξ. It follows then from (9) and (6)

$$\begin{aligned}
\sum_{\mu,\nu=0}^{n} a_{\mu\nu} z_\mu z_\nu &= \sum_{\mu=0}^{n} z_\mu \left(\sum_{\nu=1}^{n} a_{\mu\nu} z_\nu + a_{\mu 0} z_0 \right) \\
&= \sum_{\mu=0}^{n} z_\mu \sum_{\nu=1}^{n} a_{\mu\nu} x_\nu = \sum_{\nu=1}^{n} x_\nu \sum_{\mu=0}^{n} a_{\mu\nu} z_\mu \\
&= \sum_{\nu=1}^{n} x_\nu \left(\sum_{\mu=1}^{n} a_{\mu\nu} z_\mu + a_{0\nu} z_0 \right) \\
&= \sum_{\nu=1}^{n} x_\nu \sum_{\mu=1}^{n} a_{\mu\nu} x_\mu,
\end{aligned}$$

$$\sum_{\mu,\nu=0}^{n} a_{\mu\nu} z_\mu z_\nu = \sum_{\mu,\nu=1}^{n} a_{\mu\nu} x_\mu x_\nu. \tag{13}$$

7. It is obvious that the algebraic identity corresponding to (7) remains true for the system (10), although the corresponding quadratic form is only *semi-definite*. Therefore from (13) it follows that the relation (7) is also true for $N_k = 0$ where $r_0^{(k)}$ is given by (12), and the quadratic form K is the positive definite quadratic form (6). But then it follows from (7) that in any case

$$\lim_{k\to\infty} (r_{N_k}^{(k)})^2 / A_{N_k} = 0. \tag{14}$$

8. We discuss first the cyclic one step iteration. In this case we will prove that the procedure remains convergent, but for any $n \geqslant 2$ there exist matrices for which the modified procedure is slower and others for which the modified procedure is indeed faster than the original one. This agrees with results mentioned by Forsythe and Motzkin [3], footnote 24.

9. It follows from (14) that $\lim r_{N_k}^{(k)} = 0$. Therefore from (5) and the corresponding formula for ζ_k with $N_k = 0$,

$$x_\mu^{(k+1)} - x_\mu^{(k)} \to 0 \quad (k \to \infty\,;\ \mu = 1, \cdots, n),$$

and by (4)

$$r_\mu^{(k+1)} - r_\mu^{(k)} \to 0 \quad (k \to \infty\,;\ \mu = 1, \cdots, n).$$

More generally for each constant integer γ,

$$r_\mu^{(k+\gamma)} - r_\mu^{(k)} \to 0 \quad (k \to \infty\,;\ \mu = 1, \cdots, n).$$

But for any fixed μ, among the $n + 1$ consecutive values of k there is one for which $N_k = \mu$; therefore it follows that

$$r_\mu^{(k)} \to 0 \quad (k \to \infty\,;\ \mu = 1, \cdots, n),$$

and, since the determinant in (4) does not vanish,

$$x_\mu^{(k)} \to 0 \quad (k \to \infty\,;\ \mu = 1, \cdots, n).$$

We see that the *modified procedure* is indeed *convergent.*

10. In comparing the rate of convergence of the original and the modified procedure it is better to change our notations in the following way. If we start with a vector $\xi_0 = X_0$ and apply the complete n-cycle of single steps corresponding to $N_k = 1, \cdots, n$, the obtained vector will be denoted by X_1, and the vectors obtained in repeating each time the complete n-cycle will be denoted by X_2, $X_3, \cdots, X_k = \xi_{nk}, \cdots$.

In the same way, in the modified cyclic procedure we obtain, starting from a vector $\zeta_0 = Z_0$ and applying each time the *whole* $(n + 1)$*-cycle* corresponding to $N_k = 0, 1, \cdots, n$, the sequence of vectors $Z_1, Z_2, \cdots, Z_k = \zeta_{nk}, \cdots$.

11. We decompose A in the following way:

$$A = L + D + L^*, \tag{15}$$

where D is the diagonal matrix with the elements $a_{11}, \cdots, a_{nn}$, while in L all elements to the right and on the main diagonal and in L^* all elements to the left and on the main diagonal vanish. Then the rate of convergence of the usual cyclic single step iteration depends on the maximum modulus λ_N of the roots of the equation

$$|\lambda(L + D) + L^*| = 0, \tag{16}$$

and we have, if $\lambda_N > 0$:

$$X_k = O(\lambda_N{}^k k^{n-2}) \quad (k \to \infty), \tag{17}$$

while the starting vector X_0 can be chosen so that $X_k \lambda_N{}^{-k}$ does not tend to 0 as $k \to \infty$. The proof of this is quite similar to the following proof of the corresponding results for the modified single step iteration; see (25). If $\lambda_N = 0$, then X_1 vanishes identically, and the solution is obtained at the most in n steps.

12. We will now characterize in a similar way the rate of convergence of the

modified procedure. We have for the matrix $\hat{A}$ of the system (10) the decomposition corresponding to (15):

$$\hat{A} = \begin{pmatrix} a_{00} & a_{0\nu} \\ a_{\nu 0} & A \end{pmatrix} = \hat{L} + \hat{D} + \hat{L}^* \tag{18}$$

and obtain between Z_0 and Z_1, as in the theory of the usual cyclic one step iteration, the relations

$$\begin{aligned} &(\hat{L} + \hat{D})Z_1 + \hat{L}^* Z_0 = 0, \\ &Z_1 = -(\hat{L} + \hat{D})^{-1}\hat{L}^* Z_0. \end{aligned} \tag{19}$$

13. As has been mentioned above the result of this operation is not changed if Z_0 is replaced by the corresponding reduced vector $\hat{Z}_0 = (0, x_1^{(0)}, \cdots, x_n^{(0)})$. Before we go on from Z_1, we replace therefore Z_1 again by the corresponding reduced vector $\hat{Z}_1 = (0, x_1^{(1)}, \cdots, x_n^{(1)})$. For this purpose we apply the transformation $x_\mu = Z_\mu - Z_0$ $(\mu = 1, \cdots, n)$ which is equivalent to left multiplication by the $(n + 1) \times (n + 1)$-matrix

$$N_0 = \begin{pmatrix} 0 & 0 \\ -1 & E_n \end{pmatrix}, \tag{20}$$

where the first row consists of 0's and the first column, with the exception of the 0 on the top, of -1's.

We have then finally, putting

$$Q_0 = -N_0(\hat{L} + \hat{D})^{-1}\hat{L}^*, \tag{21}$$

$$\hat{Z}_k = Q_0^k \hat{Z}_0 \quad (k = 1, 2, \cdots). \tag{22}$$

14. If Q_0 vanishes identically, we have at once $Z_1 = \cdots = Z_k = \cdots = 0$, and the solution of (1) is attained at the first step. We can and will therefore assume that $Q_0 \neq 0$. We use then the following result due to Werner Gautschi [5].

If for any matrix $C = (c_{\mu\nu})$ we define as its "norm"

$$N(C) = \sqrt{\sum_{\mu,\nu} |c_{\mu\nu}|^2},$$

then if C is a square matrix of order n for which the greatest modulus of a fundamental root is Λ, we have

$$N(C^k) = O(\Lambda^k k^{p-1}) \quad (k \to \infty), \tag{23}$$

where p is the greatest multiplicity of a fundamental root of C with the modulus Λ. As a matter of fact (23) is not the *best* possible result, cf. [10], p. 5, Satz V. But Werner Gautschi's result is completely sufficient for our purpose.

15. If we apply this to the singular matrix (21) and denote the maximal modulus of a fundamental root of Q_0 by λ_G, then p does not exceed $n - 1$ if $\lambda_G > 0$, as will follow later from (31). We have therefore

$$N(Q_0^k) = O(\lambda_G^k k^{n-2}) \quad (k \to \infty). \tag{24}$$

Further it follows from (22), in applying the Cauchy-Schwarz inequality,

$$|\hat{Z}_k| \leqslant N(Q_0^k)\,|Z_0|,$$

and we obtain therefore

$$\hat{Z}_k = O(\lambda_G^k k^{n-2}) \quad (k \to \infty). \tag{25}$$

16. On the other hand it is easy to show that for a conveniently chosen starting vector Z_0 the expression $Z_k\lambda_G^{-k}$ does not tend to zero. Indeed, if η is an eigenvector of Q_0 corresponding to a fundamental root λ with $|\lambda| = \lambda_G$, we have

$$\lambda\eta' = Q_0\eta'$$

and iterating

$$\lambda^k\eta' = Q_0^k\eta'.$$

But, since the first row in Q_0 consists of zeros, the vector η is a reduced one and can be taken as $\hat{Z}_0$. Then we have

$$\hat{Z}_k = \lambda^k\hat{Z}_0, \quad \hat{Z}_k\lambda_G^{-k} = \left(\frac{\lambda}{\lambda_G}\right)^k \hat{Z}_0,$$

and this does not tend to zero as $k \to \infty$. If $\lambda_G = 0$, then Z_1 vanishes identically.

17. We shall now transform the fundamental equation of Q_0, and we introduce for this purpose the matrix

$$N_\epsilon = \begin{pmatrix} \epsilon & 0 \\ -1 & E_n \end{pmatrix}, \tag{26}$$

where the first row and the first column consist respectively of 0's and -1's with the exception of the first element ϵ. N_ϵ corresponds to the transformation

$$y_0 = \epsilon Z_0, \quad y_\nu = Z_\nu - Z_0 \quad (\nu = 1, \cdots, n) \tag{27}$$

and for $\epsilon \to 0$ goes into N_0. Since the inverse of (27) is for $\epsilon \neq 0$

$$Z_0 = \frac{1}{\epsilon}y_0, \quad Z_\nu = y_\nu + \frac{1}{\epsilon}y_0 \quad (\nu = 1, \cdots, n),$$

we have

$$N_\epsilon^{-1} = \begin{pmatrix} \epsilon^{-1} & 0 \\ \epsilon^{-1} & E_n \end{pmatrix}, \tag{28}$$

where the first column consists of ϵ^{-1}, while the first row with the exception of the first element contains only 0's.

The fundamental equation of Q_0 can be written in the form

$$\lim_{\epsilon \to 0} |\lambda E + N_\epsilon(\hat{L} + \hat{D})^{-1}\hat{L}^*| = 0. \tag{29}$$

18. On the other hand we have identically, since $|N_\epsilon| = \epsilon$,

$$(|\hat{L} + \hat{D}|)(|\lambda E + N_\epsilon(\hat{L} + \hat{D})^{-1}\hat{L}^*|) = \epsilon|\lambda(\hat{L} + \hat{D})N_\epsilon^{-1} + \hat{L}^*|, \tag{30}$$

and we obtain the fundamental equation of Q_0 in taking the limit for $\epsilon \to 0$ on the right in (30).

Now we have

$$(\hat{L} + \hat{D})N_\epsilon^{-1} = \begin{pmatrix} \epsilon^{-1}a_{00} & 0 \\ \epsilon^{-1}\sum_{\mu \le \nu} a_{\nu\mu} & L + D \end{pmatrix},$$

where $a_{n0} + a_{n1} + \cdots + a_{nn} = 0$ by (9); and we have therefore identically

$$\epsilon|\lambda(\hat{L} + \hat{D}) + \hat{L}^*| = \lambda \begin{vmatrix} a_{00} & a_{0\nu} \\ \sum_\mu a_{\nu\mu} & \lambda(L + D) + L^* \end{vmatrix}.$$

λ_G is therefore the maximum modulus of the roots of the equation of degree n

$$G(\lambda) = \begin{vmatrix} a_{00} & a_{0\nu} \\ \sum_{\mu \le \nu} a_{\nu\mu} & \lambda(L + D) + L^* \end{vmatrix} = 0. \tag{31}$$

One root of this equation is zero.

19. In specializing for $n = 2$ we obtain in particular, if we put $a_{11} = a_1$, $a_{22} = a_2$, $a_{12} = a_{21} = \sigma$ and assume $\sigma \neq 0$:

$$G_2(\lambda) = \begin{vmatrix} a_1 + a_2 + 2\sigma & -a_1 - \sigma & -a_2 - \sigma \\ -\sigma & \lambda a_1 & \sigma \\ 0 & \lambda\sigma & \lambda a_2 \end{vmatrix}$$

$$= (a_1 + a_2 + 2\sigma)a_2a_1\lambda^2 - \sigma\lambda(a_1 + \sigma)(a_2 + \sigma),$$

$$\lambda_G = \frac{|\sigma(a_1 + \sigma)(a_2 + \sigma)|}{a_1a_2(a_1 + a_2 + 2\sigma)}, \tag{32}$$

while the equation (16) for λ_N reduces to $\lambda^2 a_1a_2 - \lambda\sigma^2 = 0$, and gives

$$\lambda_N = \frac{\sigma^2}{a_1a_2}. \tag{33}$$

From (32) and (33) we have

$$\frac{\lambda_G}{\lambda_N} = \frac{|a_1 + \sigma|\,|a_2 + \sigma|}{|\sigma|\,(a_1 + a_2 + 2\sigma)}. \tag{34}$$

20. If we square this, subtract 1 and multiply by the square of the denominator we obtain

$$[(a_1 + \sigma)(a_2 + \sigma) - \sigma(a_1 + a_2) - 2\sigma^2][(a_1 + \sigma)(a_2 + \sigma) + \sigma(a_1 + a_2) + 2\sigma^2]$$
$$= (a_1a_2 - \sigma^2)(a_1a_2 + 2(a_1 + a_2)\sigma + 3\sigma^2).$$

Since the first factor is positive, we see that $\lambda_G \gtreqless \lambda_N$ according as

$$\varphi(\sigma) \equiv 3\sigma^2 + 2(a_1 + a_2)\sigma + a_1a_2 \gtreqless 0. \tag{35}$$

Here $\sigma \neq 0$ is subject only to the condition $|\sigma| \leqslant \sqrt{a_1a_2}$. We have obviously

$$\varphi(-\sqrt{a_1a_2}) = 4\sqrt{a_1a_2}\left(\sqrt{a_1a_2} - \frac{a_1 + a_2}{2}\right).$$

Since this is $\leqslant 0$, we see that $\varphi(-\sigma)$ has two positive roots σ_1, σ_2 with $\sigma_1\sigma_2 = \dfrac{a_1a_2}{3}$, and

$$0 < \sigma_1 < \sqrt{a_1a_2} \leqslant \sigma_2.$$

It follows that for $n = 2$ we have $\lambda_G < \lambda_N$, if $\sigma_1 < -\sigma < \sqrt{a_1a_2}$, and $\lambda_G > \lambda_N$, if $-\sqrt{a_1a_2} < -\sigma < \sigma_1$. The modified procedure is in the first case faster and in the second slower than the original one.

21. To prove the corresponding result for $n > 2$ consider the matrix A of the quadratic form

$$K(\xi) = a_1x_1^2 + 2\sigma x_1x_2 + a_2x_2^2 + \sum_{\mu=3}^{n} x_\mu^2. \tag{36}$$

In the corresponding determinant (31) for $G(\lambda)$ the elements in the first column are

$$a_{\mu 0} + a_{\mu 1} + \cdots + a_{\mu\mu} = -(a_{\mu,\mu+1} + \cdots + a_{\mu n})$$

and vanish therefore for $\mu \geqslant 2$. The same is true for the elements $\lambda a_{\mu\nu}$ to the left of the main diagonal with $\mu > 2$ and $\nu < \mu$, while the elements on the diagonal, $\lambda a_{\mu\mu}$ $(\mu > 2)$, become respectively λ. We obtain therefore

$$G(\lambda) = \lambda^{n-2}G_2(\lambda)$$

so that λ_G is given in this case by (32).

22. In the same way it follows from (16) that λ_N is again given by (33). We can have in this case, according to the chosen values of σ, either $\lambda_G > \lambda_N$ or $\lambda_G < \lambda_N$.

It may finally be remarked that the value of λ_G is not changed, if the $(n+1)$-st equation in (10) and the corresponding new variable z_0 are not put at the beginning but are interpolated between two indices μ, $\mu + 1$ or even put at the end. Indeed this amounts to the old process applied to a transformation of $\hat{\zeta}_0$ by a finite sequence of single step iterations, but then $\hat{\zeta}_0$ is carried over into the *general* reduced vector, and the invariance of λ_G follows then from the characterization of λ_G contained in the developments of numbers 15 and 16.

23. We consider now Seidel's relaxation procedure (8). Then in the modified procedure we obtain the speeding up for an index k for which we have

$$(r_0^{(k)})^2/A_0 > (r_\mu^{(k)})^2/A_\mu \quad (\mu \neq 0), \tag{37}$$

that is to say, by (12):

$$|\sum_{\nu=1}^{n} r_\nu^{(k)}|/A_0^{\frac{1}{2}} > |r_\mu^{(k)}|/A_\mu^{\frac{1}{2}} (\mu \neq 0). \tag{38}$$

We will now show that this inequality is impossible, if we have

$$A_0^{\frac{1}{2}} \geq \sum_{\nu=1}^{n} A_\nu^{\frac{1}{2}} - \min_{1 \leqslant \mu \leqslant n} A_\mu^{\frac{1}{2}}. \tag{39}$$

Indeed, if we put

$$|r_\mu^{(k)}| = p_\mu A_\mu^{\frac{1}{2}}, \quad \mu = 1, \cdots, n,$$

we have from (38) and (39), since one of the r_μ and therefore one of the p_μ vanish:

$$|\sum_{\nu=1}^{n} r_\nu^{(k)}|/A_0^{\frac{1}{2}} \leqslant \sum_{\nu=1}^{n} p_\nu (A_\nu/A_0)^{\frac{1}{2}} \leqslant A_0^{-\frac{1}{2}}(-\min_{1\leqslant\mu\leqslant n} A_\mu^{\frac{1}{2}} + \sum_{\nu=1}^{n} A_\nu^{\frac{1}{2}}) \max p_\mu$$

$$\leqslant \max p_\mu = \max_{1\leqslant\mu\leqslant n} (|r_\mu^{(k)}| A_\mu^{-\frac{1}{2}}).$$

Therefore, in order that (37) be possible at all we must have

$$\sqrt{A_0} < \sum_{\mu=1}^{n} \sqrt{A_\mu} - \min_{1\leqslant\mu\leqslant n} \sqrt{A_\mu}. \tag{40}$$

24. In order to discuss the situation under the condition (40) we consider the $r_\mu^{(k)}$ as *stochastic variables* and discuss the probability for (37) under suitable assumptions on the distribution of the $r_\mu^{(k)}$. We put

$$\beta_\nu = \sqrt{A_\nu} > 0 \quad (\nu = 0, 1, \cdots, n) \tag{41}$$

and denote the sequence of the vectors $\rho_k (k = 0, 1, \cdots)$ by S. In the following formulae the index of ρ, that is the upper index of $r_\nu^{(k)}$, will be dropped whenever it is possible without danger of misunderstanding.

Then our problem can be described as the problem of computing the probability

$$P\left[\frac{1}{\beta_0}\left|\sum_{\nu=1}^{n} r_\nu\right| > \max_{1\leqslant\nu\leqslant n} \frac{|r_\nu|}{\beta_\nu}; \quad \rho \in S\right] \tag{42}$$

defined in the usual way as the limit of the relative density, and ascertaining whether (42) is positive. As the relations in the brackets of (42) are homogeneous in the r_ν, each of the vectors $\rho \in S$ can and is from now on assumed to be normed in such a way that we have

$$\max_{1\leqslant\nu\leqslant n} \frac{|r_\nu|}{\beta_\nu} = 1. \tag{43}$$

25. Denote the N-th section $(\rho_1, \cdots, \rho_N)$ of S by $S^{(N)}$ and put

$$\Phi^{(N)}(\sigma) = P\left[\sum_{\nu=1}^{n} r_\nu \leqslant \sigma, \quad \rho \in S^{(N)}\right]. \tag{44}$$

This is the "finite" probability in the classical sense, and we put then

$$\Phi(\sigma) = \lim_{N\to\infty} \Phi^{(N)}(\sigma), \tag{45}$$

if this limit exists.

For any k, $k = 1, \cdots, n$, we denote by $S_k^{(N)}$ the sequence of the vectors ρ_λ, $\lambda \leqslant N$, for which k is leading index corresponding to $\rho_{\lambda-1}$ and therefore $r_k^{(\lambda)} = 0$.

Further we write S_k for the complete partial sequence of S corresponding to these vectors.

We denote further by $L_\mu(N)$ the number of vectors ρ_λ in $S^{(N)}$ whose leading index is μ and put

$$\Phi_k^{(N)}(\sigma) = P\left[\sum_{\nu=1}^{n} r_\nu \leqslant \sigma; \quad \rho \in S_k^{(N)}\right]. \tag{46}$$

Denote now by $S_{k,\mu}$ the partial sequence of S containing the vectors ρ_λ such that μ is leading index for ρ_λ and k is leading index for $\rho_{\lambda-1}$, and therefore by (8) and (43) $r_k^{(\lambda)} = 0$, $|r_\mu^{(\lambda)}| = \beta_\mu$, and by $S_{k,\mu}^{(N)}$ the sequence of the vectors ρ_λ from $S_{k,\mu}$ with $\lambda \leqslant N$. Put then

$$\Phi_{k,\mu}^{(N)}(\sigma) = P\left[\sum_{\nu=1}^{n} r_\nu \leqslant \sigma; \rho \in S_{k,\mu}^{(N)}\right], \tag{47}$$

and denote by $H_{k,\mu}(N)$ the probability that the leading index μ in $S^{(N)}$ follows the leading index k.

26. We define further

$$H_k = \lim_{N\to\infty} \frac{1}{N} L_k(N), \tag{48}$$

$$H_{k,\mu} = \lim_{N\to\infty} H_{k,\mu}(N), \tag{49}$$

$$\Phi_k(\sigma) = \lim_{N\to\infty} \Phi_k^{(N)}(\sigma), \tag{50}$$

$$\Phi_{k,\mu}(\sigma) = \lim_{N\to\infty} \Phi_{k,\mu}^{(N)}(\sigma) \tag{51}$$

if the limits (48)–(51) exist. We have then obviously by elementary probabilities

$$\Phi^{(N)}(\sigma) = \sum_{k=1}^{n} \Phi_k^{(N)}(\sigma) \frac{L_k(N)}{N+1}, \tag{52}$$

$$\Phi_k^{(N)}(\sigma) = \sum_{\mu=1}^{n} \Phi_{k,\mu}^{(N)}(\sigma) H_{k,\mu}(N). \tag{53}$$

We see that, if the limits (48), (49) and (51) exist, $\Phi_k(\sigma)$ and $\Phi(\sigma)$ also exist and we have

$$\Phi(\sigma) = \sum_{k=1}^{n} \Phi_k(\sigma) H_k, \tag{54}$$

$$\Phi_k(\sigma) = \sum_{\mu=1}^{n} \Phi_{k,\mu}(\sigma) H_{k,\mu}. \tag{55}$$

27. We define now n vector-fields F_k consisting respectively of all vectors ρ normed according to (43), for which the component r_k vanishes. Each of the fields

F_k can again be decomposed into $n-1$ vector-fields $F_{k,\mu}(k \neq \mu)$ consisting respectively of all vectors ρ normed according to (43) with $r_k = 0$, $|r_\mu| = \beta_\mu$. A vector ρ can of course belong to several of these fields.

In the fields F_k, $F_{k,\mu}$ we will now define in a suitable way the following *geometric probabilities:*

$$\Phi_k^*(\sigma) = P\left[\sum_{\nu=1}^{n} r_\nu \leqslant \sigma;\ \rho \in F_k\right], \tag{56}$$

$$\omega_{k,\mu} = P[|r_\mu| = \beta_\mu;\ \rho \in F_k], \tag{57}$$

$$\Phi^*_{k,\mu}(\sigma) = P\left[\sum_{\nu=1}^{n} r_\nu \leqslant \sigma;\ \rho \in F_{k,\mu}\right]. \tag{58}$$

Of course, to define these geometric probabilities we must use a convenient model, which presents itself here in a very natural way. If the components $r_\nu^{(\lambda)}$ of the residual vector ρ_λ are computed up to a certain number of decimal places, they are approximated by certain fractions $\frac{q_\nu^{(\lambda)}}{m}$, where m is a power of 10. It is then natural to assume that, if the subscripts k and μ are fixed, the integers $q_\nu^{(\lambda)}$ vary uniformly within their limits. We consider therefore $n+1$ integers

$$m, m_\nu \quad (\nu = 1, \cdots, n)$$

tending to infinity in such a way that we have

$$\frac{m_\nu}{m} \to \beta_\nu \quad (\nu = 1, \cdots, n).$$

We define then

$$\Phi_k^*(\sigma) = \lim_{m\to\infty} P\left[\sum_{\nu=1}^{n} q_\nu \leqslant \sigma m;\ \max_{1\leqslant\nu\leqslant n} \frac{|q_\nu|}{m_\nu} = 1,\ q_k = 0\right], \tag{59}$$

$$\omega_{k,\mu} = \lim_{m\to\infty} P\left[|q_\mu| = m_\mu;\ \max_{1\leqslant\nu\leqslant n} \frac{|q_\nu|}{m_\nu} = 1,\ q_k = 0\right], \tag{60}$$

$$\Phi^*_{k,\mu} = \lim_{m\to\infty} P\left[\sum_{\nu=1}^{n} q_\nu \leqslant \sigma m;\ |q_\mu| = m_\mu,\ q_k = 0,\ \max_{1\leqslant\nu\leqslant n} \frac{|q_\nu|}{m_\nu} = 1\right]. \tag{61}$$

28. It is easy to compute the $\omega_{k,\mu}$. Assume $k = 1$. Then the probability $\omega_{1,\mu}$ in (60) is obtained in considering the right parallelepiped $|x_\lambda| \leqslant m_\lambda$ $(\lambda = 2, \cdots, n)$, as the quotient of the area of the two faces $|x_\mu| = m_\mu$ by the complete area of all faces. In putting temporarily $m_2 \cdots m_n = M$ we obtain

$$\frac{2M/m_\mu}{2\sum_{\lambda=2}^{n} M/m_\lambda} = \frac{m/m_\mu}{\sum_{\lambda=2}^{n} m/m_\lambda},$$

and this tends to

$$\frac{1/\beta_\mu}{\sum_{\nu=1}^{n} 1/\beta_\nu - 1/\beta_1}.$$

In replacing here β_1 by β_k we obtain finally

$$\omega_{k,\mu} = \frac{1/\beta_\mu}{\sum_{\nu=1}^{n} 1/\beta_\nu - 1/\beta_k} \quad (k \neq \mu). \tag{62}$$

We write this differently in introducing the expressions

$$\sigma_\lambda = \frac{1/\beta_\lambda}{\sum_{\nu=1}^{n} 1/\beta_\nu}. \tag{63}$$

Then we have obviously

$$\sum_{\lambda=1}^{n} \sigma_\lambda = 1, \tag{64}$$

and (62) becomes

$$\frac{\sigma_\mu}{\sum_{\nu=1}^{n} \sigma_\nu - \sigma_k} = \frac{\sigma_\mu}{1 - \sigma_k}.$$

If we now define $\omega_{k,k}$ as 0, we obtain

$$\omega_{k,\mu} = \begin{cases} \dfrac{\sigma_\mu}{1 - \sigma_k} & (k \neq \mu), \\ 0 & (k = \mu). \end{cases} \tag{65}$$

The values $\Phi^*_{k,\mu}(\sigma)$ and $\Phi_k^*(\sigma)$ have already been computed in another communication [8], and we will give $\Phi^*_{k,\mu}(\sigma)$ later on. In particular we have

$$\Phi_k^*(\sigma) = \sum_{\substack{\mu=1 \\ \mu \neq k}}^{n} \omega_{k,\mu} \Phi^*_{k,\mu}(\sigma). \tag{66}$$

29. We make now two fundamental assumptions about the sequence S under consideration:

(67) *Hypothesis A.* $\Phi_{k,\mu}(\sigma) = \Phi^*_{k,\mu}(\sigma)$ $(k \neq \mu;\ k, \mu = 1, \cdots, n)$

(68) *Hypothesis B.* $H_{k,\mu}(N) \to \omega_{k,\mu}$ $(N \to \infty,\ k \neq \mu;\ k, \mu = 1, \cdots, n)$.

It follows then from (55) that we have

$$\Phi_k(\sigma) = \Phi_k^*(\sigma). \tag{69}$$

In order to obtain the value of $\Phi(\sigma)$ we will now prove the existence of the limit (48).

30. We have obviously

$$\sum_{k=1}^{n} L_k(N) = N + 1, \tag{70}$$

$$L_\mu(N) = \sum_{\substack{k=1 \\ k \neq \mu}}^{n} H_{k,\mu}(N) L_k(N-1).$$

If we replace here each $L_k(N-1)$ by $L_k(N)$, the modulus of the error does not exceed $\sum_{\substack{k=1 \\ k \neq \mu}}^{n} H_{k,\mu} = 1$, and we obtain

$$\sum_{\substack{k=1 \\ k \neq \mu}}^{n} H_{k,\mu}(N) L_k(N) = L_\mu(N) + e_\mu(N), \quad |e_\mu(N)| \leqslant 1.$$

Divide this by N and put

$$\frac{L_k(N)}{N} = h_k(N). \tag{71}$$

Then we obtain

$$\sum_{\substack{k=1 \\ k \neq \mu}}^{n} H_{k,\mu}(N) h_k(N) - h_\mu(N) = \epsilon_\mu(N), \quad \epsilon_\mu(N) = O\left(\frac{1}{N}\right) \quad (\mu = 1, \cdots, n). \tag{72}$$

In the same way we obtain from (70)

$$\sum_{k=1}^{n} h_k(N) = 1 + \frac{1}{N}. \tag{73}$$

If we replace in the equations (72) and (73) the $h_\mu(N)$ by x_μ, the equations become for $N \to \infty$ by (68) and (65)

$$\sigma_\mu \sum_{\substack{k=1 \\ k \neq \mu}}^{n} \frac{x_k}{1 - \sigma_k} = x_\mu \quad (\mu = 1, \cdots, n), \tag{74}$$

$$\sum_{k=1}^{n} x_k = 1. \tag{75}$$

We consider first the system (74). If we divide by σ_μ and add $\dfrac{x_\mu}{1 - \sigma_\mu}$ on both sides, we obtain

$$\sum_{k=1}^{n} \frac{x_k}{1 - \sigma_k} = \frac{x_\mu}{\sigma_\mu} + \frac{x_\mu}{1 - \sigma_\mu} = \frac{x_\mu}{\sigma_\mu(1 - \sigma_\mu)}.$$

Denoting the expression on the left by γ we get

$$x_\mu = \gamma \sigma_\mu (1 - \sigma_\mu). \tag{76}$$

31. On the other hand, if we put the value (76) for $\gamma = 1$ into the system (74), the system is satisfied; we see that the rank of (74) is exactly $n - 1$. We can

therefore choose $n-1$ of the equations (74) in such a way that they form, taken together with (75), a system of the rank n. Therefore the corresponding $n-1$ equations (72) form, taken together with the equation (73), a linear system which remains regular in the limit $N \to \infty$. Therefore its solutions $h_k(N)$ tend to the solution of the system (74), (75); and this solution is obtained from (76), if γ is chosen so that (75) is fulfilled. But then we have

$$\sum_{\mu=1}^{n} x_\mu = 1 = \gamma\left(1 - \sum_{\mu=1}^{n} \sigma_\mu^2\right),$$

$$\frac{L_k(N)}{N} \to H_k = \frac{\sigma_k(1-\sigma_k)}{1 - \sum_{\nu=1}^{n} \sigma_\nu^2}. \tag{77}$$

If we introduce here the values (63) of the σ_ν, we obtain

$$H_k = \frac{\beta_k^{-1}(\sum_\lambda \beta_\lambda^{-1} - \beta_k^{-1})}{2 \sum_{\nu \neq \lambda} \beta_\nu^{-1}\beta_\lambda^{-1}}. \tag{78}$$

It follows now from (54), (55), (68), (65):

$$\Phi(\sigma) = \sum_{k=1}^{n} H_k \sum_{\substack{\mu=1 \\ \mu \neq k}}^{n} \omega_{k,\mu}\Phi^*_{k,\mu}(\sigma). \tag{79}$$

32. An explicit formula for $\Phi^*_{k,\mu}(\sigma)$ can be written simply by using the expression for a function $F(\sigma)$ which was defined and computed in a previous paper [8]. In replacing the integer n used in that paper by $n-2$ (for $n \geqslant 3$) we define $F(\sigma)$ as the probability

$$F(\sigma) = P[\xi_1 + \xi_2 + \cdots + \xi_{n-2} \leqslant \sigma;\ |\xi_\nu| \leqslant \alpha_\nu \quad (\nu = 1, \cdots, n-2)]. \tag{80}$$

The α_ν are $n-2$ positive numbers. Then we have for $F(\sigma)$ (see formula (5) in [8]):

$$F(\sigma) = \frac{1}{2^{n-2}(n-2)!} \cdot \frac{1}{\alpha_1 \cdots \alpha_{n-2}} \prod_{\nu=1}^{n-2} (1 - S^{2\alpha_\nu})(\alpha + \sigma)_+^{n-2}, \tag{81}$$

$$\alpha = \alpha_1 + \cdots + \alpha_{n-2}.$$

Here S is the operator defined by

$$S^\eta f(\alpha) = f(\alpha - \eta).$$

The computed expression consists of 2^{n-2} terms of the form

$$\pm\,(\alpha - 2\alpha_{\lambda_1} - 2\alpha_{\lambda_2} - \cdots + \sigma)_+^{n-2}.$$

The subscript $+$ signifies that, if the expression within the parentheses is negative, the whole expression has to be replaced by 0, while otherwise the subscript can be dropped.

In particular we have pointed out in [8], p. 6, that $F(\sigma)$ is *strictly monotonically increasing* with σ for $-\alpha \leqslant \sigma \leqslant \alpha$ and grows from $F(-\alpha) = 0$ to $F(\alpha) = 1$.

We see that we have

$$1 > F(\sigma) > 0 \quad (|\sigma| < \alpha). \tag{82}$$

$F(\sigma)$ is further a continuous function of σ. We obtain therefore the same value of the probability, if we replace in (80) the condition $\leqslant \sigma$ by the condition $< \sigma$.

33. $\Phi^*_{k,\mu}(\sigma)$ is by definition (58) the arithmetical mean of the two probabilities

$$P\left[\sum_{\substack{\nu=1\\ \nu\neq k,\mu}}^{n} r_\nu \leqslant \sigma - \beta_\mu;\ |r_\nu| \leqslant \beta_\nu\right], P\left[\sum_{\substack{\nu=1\\ \nu\neq k,\mu}}^{n} r_\nu \leqslant \sigma + \beta_\mu;\ |r_\nu| \leqslant \beta_\nu\right].$$

We have therefore

$$\Phi_{k,\mu}(\sigma) = \tfrac{1}{2}[F_{k,\mu}(\sigma - \beta_\mu) + F_{k,\mu}(\sigma + \beta_\mu)]. \tag{83}$$

The function $F_{k,\mu}$ is here obtained from (81), if the $\alpha_1, \cdots, \alpha_{n-2}$ are identified with the $n - 2$ of the numbers β_ν which remain after deleting β_μ and β_k. Introducing (83) into (79) we obtain finally the complete expression of $\Phi(\sigma)$.

If we return now to the problem of computing the probability (42) and norm here the r_ν according to (43), we see that we have to compute

$$P\left[\left|\sum_{\nu=1}^{n} r_\nu\right| > \beta_0,\ \max_{1\leqslant\nu\leqslant n} \frac{|r_\nu|}{\beta_\nu} = 1,\ \min_{1\leqslant\nu\leqslant n} |r_\nu| = 0\right].$$

This is the sum of the two probabilities

$$P\left[\sum_{\nu=1}^{n} r_\nu < -\beta_0;\ \max_{1\leqslant\nu\leqslant n} \frac{|r_\nu|}{\beta_\nu} = 1,\ \min_{1\leqslant\nu\leqslant n} |r_\nu| = 0\right],$$

$$P\left[\sum_{\nu=1}^{n} r_\nu > \beta_0;\ \max_{1\leqslant\nu\leqslant n} \frac{|r_\nu|}{\beta_\nu} = 1,\ \min_{1\leqslant\nu\leqslant n} |r_\nu| = 0\right].$$

But these two probabilities have the same value, and the first of them is obviously $\Phi(-\beta_0)$. We obtain therefore for the probability (42) the expression

$$2\Phi(-\beta_0) = \sum_{k=1}^{n} H_k \sum_{\substack{\mu=1\\ \mu\neq k}}^{n} \omega_{k,\mu}[F_{k,\mu}(-\beta_0 - \beta_\mu) + F_{k,\mu}(-\beta_0 + \beta_\mu)]. \tag{84}$$

34. We will now prove that (84) is *positive* under the condition (40), that is

$$\beta_0 < \sum_{\nu=1}^{n} \beta_\nu - \min_{1\leqslant\nu\leqslant n} \beta_\nu. \tag{85}$$

Observe that all terms in (84) are non-negative. It is therefore sufficient to prove that under the condition (85) at least one of the expressions $F_{k,\mu}(-\beta_0 \pm \beta_\mu)$ is positive, that is to say, by (82), that for convenient choice of k and μ either $\beta_0 + \beta_\mu$ or $|\beta_0 - \beta_\mu|$ is less than $\sum_{\nu=1}^{n} \beta_\nu - \beta_\mu - \beta_k$. Assume now that the β_ν are ordered increasingly

$$\beta_1 \leqslant \beta_2 \leqslant \cdots \leqslant \beta_n$$

and put

$$\beta = \beta_1 + \cdots + \beta_n.$$

Then I say that for $k = 1$, $\mu = 2$ we always have

$$|\beta_0 - \beta_2| < \beta - \beta_1 - \beta_2, \tag{86}$$

that is to say that

$$F_{1,2}(-\beta_0 + \beta_2) > 0.$$

To prove (86) consider the two cases $\beta_0 \geqslant \beta_2$ and $\beta_0 < \beta_2$. In the first case (86) reduces to

$$\beta_0 - \beta_2 < \beta - \beta_1 - \beta_2, \quad \beta_0 < \beta - \beta_1,$$

and this is exactly (85). In the second case (86) reduces to

$$\beta_2 - \beta_0 < \beta - \beta_1 - \beta_2, \quad \beta_0 > -(\beta - \beta_1 - 2\beta_2);$$

but here the expression on the right is not positive, as

$$(\beta_3 - \beta_2) + \beta_4 + \cdots + \beta_n \geqslant 0.$$

We have now proved that under the condition (40) and the assumptions of Nr. 29 the probability for the speeding up of Seidel's relaxation procedure in using Gauss' device is *positive.*

The author gratefully acknowledges discussions with T. S. Motzkin and O. Taussky-Todd.

University of Basel
Basel, Switzerland

1. R. DEDEKIND, "Gauss in seiner Vorlesung über die Methode der kleinsten Quadrate," *Festschrift zur Feier des 150-jährigen Bestehens der königlichen Gesellschaft der Wissenschaften zu Göttingen.* Berlin, 1901, p. 45–49.
2. R. DEDEKIND, *Gesammelte mathematische Werke.* Herausgegeben von R. Fricke, E. Noether, & Ö. Ore. Bd. II. 1931, Braunschweig. p. 293–306, especially p. 300–301.
3. G. E. FORSYTHE & T. S. MOTZKIN, "An extension of Gauss' Transformation for improving the condition of systems of linear equations," *MTAC*, v. 6, 1952, p. 9–17.
4. C. F. GAUSS, "Letter to Gerling," 26 December 1823, *Werke*, v. 9, p. 278–281. For an annotated translation of Gauss' letter by G. E. Forsythe, see *MTAC*, v. 5, 1951, p. 255–258.
5. WERNER GAUTSCHI, "The asymptotic behaviour of the powers of matrices," *Duke Mathematical Journal*, v. 20, 1953, p. 127–140.
6. CHRISTIAN LUDWIG GERLING, *Die Ausgleichsrechnung der practischen Geometrie.* Hamburg and Gotha, 1843.
7. F. R. HELMERT, *Die Ausgleichsrechnung nach der Methode der kleinsten Quadrate, mit Anwendung auf die Geodäsie und die Theorie der Messinstrumente*, Leipzig, 1872, p. 136.
8. A. M. OSTROWSKI, "Two explicit formulae for the distribution function of the sums of n uniformly distributed independent variables," *Archiv der Mathematik*, v. 3, 1952, p. 3–11.
9. A. M. OSTROWSKI, "On the linear iteration procedures for symmetric matrices," *Rendiconti di Matematica e delle sue applicazioni*, Serie V, v. 8, 1954, p. 140–163.
10. A. M. OSTROWSKI, "Ueber Normen von Matrizen," *Math. Zeitschr.*, v. 63, 1955, p. 2–18.
11. E. REICH, "On the convergence of the classical iterative method of solving linear simultaneous equations," *Annals Math. Stat.*, v. 20, 1949, p. 448–451.
12. W. SCHMEIDLER, *Vorträge über Determinanten und Matrizen mit Anwendungen in Physik und Technik*, Berlin, 1949.
13. LUDWIG SEIDEL, "Über ein Verfahren, die Gleichungen, auf welche die Methode der kleinsten Quadrate führt, sowie lineare Gleichungen überhaupt, durch successive Annäherung aufzulösen," Akad. Wiss., Munich, *Mathematisch-Naturwissenschaftliche Abteilung*, v. 11, No. 3, 1874, p. 81–108.
14. R. V. SOUTHWELL, "Stress-calculation in frameworks by the method of systematic relaxation of constraints," I & II, Roy. Soc. London, *Proc.*, A 151, 1935, p. 56–95.
15. R. ZURMUHL, *Matrizen, Eine Darstellung für Ingenieure.* Berlin, 1950, p. 280 ff. [See *MTAC*, v. 5, 1951, p. 13–14.]