Practical procedures for equation solving

Richard von Mises and Hilda Pollaczek-Geiringer

Zamm.-Z Aengew. Math. Me., v 9, (1929), pp. 58-77

[This is only a partial translation of the paper]

In a lecture on "Practical Analysis" given by the first-named author in the summer semester of 1927, a number of calculation methods were discussed, some of which are new and some of which are little known. It seemed appropriate to compile some of them with a short justification and references to the most important applications. This work has been undertaken by the second author - first of all with respect to the methods of solving equations; essential additions and more detailed explanations as well as the numerical examples also originate from her.

In the first section of the present report a method for the solution of an arbitrary equation with one unknown is developed, which differs from those known so far in that it always leads to the objective (always converges) and generally requires less computational work. Finding the roots of systems of equations is treated in the following two sections only for the special case of linear equations. For the "iteration method in total steps", which represents the direct transfer of the method for one unknown, in section 2, partly known, mostly not properly considered, convergence conditions and error estimates are given.

The approach from Ph. Seidel, described in Section 3 and referred to as "iteration in single steps", is of particular importance. After certain transformations, this can be applied to any linear system; on the other hand, in its original form, without any preparatory work, it provides a general, always convergent method for the solution of a comprehensive class of linear equations, namely all those that result from a minimum problem. For the treatment of n-fold statically indeterminate systems one can derive a very clear theorem from this, which reduces the general problem to the repeated solution of simply indeterminate systems[1] (Section 4).

Finally, in the last sections, systems of homogeneous linear equations with parameters are considered, a task which is known to arise in the study

[1] See the lecture excerpt: H. Pollaczek-Geiringer, this Zeitschr. vol. 8 (1928), p. 446.

of natural oscillations of elastic systems and in the search for stability limits (critical loads). The analogous transfer of the method of Vianello, known in technology, for the determination of the smallest eigenvalue in boundary value problems of ordinary differential equations allows to find the smallest eigenvalue and the corresponding eigensolution by simple iteration, without the previous solution of an equation of nth degree. That one can find the largest eigenvalues and associated eigensolutions in the same way, by repeated application of the method, is an often repeated, inaccurate assertion. In Section 6 we indicate how this problem, which is often important for applications, can actually be solved.

Where relevant earlier publications have come to our attention, they have been cited at the relevant places. The sentences and other results reproduced without supporting documents are likely to be essentially new.

1. Solving an equation with one unknown.

To find a root a of the equation $f(x) = 0$ [where $f(x)$ is a real continuous function of the real variable x] means in practice to determine an interval of a sufficiently small length δ within which f vanishes at least once, e.g. the specification that $a = 0.276$ is a root of $f(x) = 0$ means that $f(x)$ takes the value zero at least once within the interval from 0.2755 to 0.2765. In particular, the task may be considered, starting from a given point $x = x_1$, to find the next root of $f(x)$ to the right of x_1.

To determine such a root, one must first get an idea of the increase of the function, of its "slope", either by drawing a rough sketch or by estimating the difference quotient. Strictly speaking, the slope is the differential quotient

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

One now chooses a number c, which is smaller than the reciprocal value of the largest slope in the area in question for finding the root:

$$|c| \leq \max \left| \frac{x_2 - x_1}{f(x_2) - f(x_1)} \right|. \qquad (1)$$

For a differentiable function, one will usually make use of the fact that for such a function, the upper and lower limits of the slope in a given interval coincide with the upper and lower limit of the derivative in this interval. With a value of c chosen according to (1), starting from a point x_1 where $f(x_1)$ may be positive, one forms the sequence of numbers:

$$x_2 = x_1 + cf(x_1), \quad x_3 = x_2 + cf(x_2), \ldots \qquad (2)$$

We claim that in this way, if there is a root to the right or to the left of x_1 at all, one always arrives at this root, namely at the next one to the right of x_1, if c is positive, and at the left one, if c is negative. If $f(x_1) < 0$, then right and left are interchanged.

Geometrically, (2) means that, starting from the curve point $y_1 = f(x_1)$, a straight line is drawn at the angle α (Fig. 1), where $c = \cot\alpha$, i.e. a straight line which, because of (1), descends more steeply to the right than the strongest chordal slope of the curve in the interval under consideration (between x_1 and the root). From the figure - which is drawn for positive f - you can see that in this way from the left one gets closer and closer to the root without going beyond it.

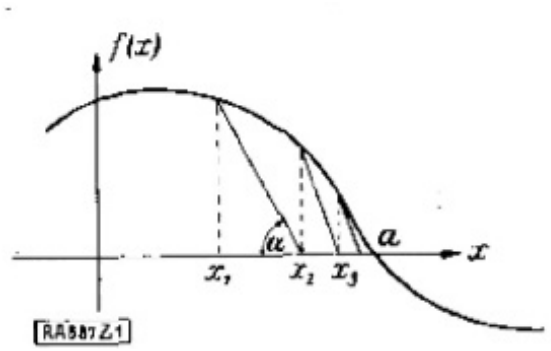

Fig. 1

For the mathematician the proof of convergence is based on the following considerations: According to the explanation on c, it is smaller than the reciprocal value of the slope for any two values; thus, if the slope is formed between the nth approximate value x_n, and a root a, then

$$|c| < \left|\frac{x_n - a}{f(x_n) - f(a)}\right| = \left|\frac{x_n - a}{f(x_n)}\right|,$$

and therefore, because of (2), for each $n = 1, 2, \ldots$

$$|x_{n+1} - x_n| = |cf(x_n)| < |x_n - a|. \qquad (3)$$

If $f(x_1) > 0$, c is chosen to be positive, and a denotes the next root to the right of x_1, then it follows from (2) that $x_2 > x_1$ and from (3), because a is to the right of x_1, $x_2 - x_1 < a - x_1$ i.e. x_2, lies between x_1 and a. Accordingly, since a is the next root to the right of x_1, $f(x_2)$ must also be > 0, and we conclude again from (2) and (3) that x_3, must lie between x_2, and a, and so on. The $x_1, x_2, \ldots$ form a bounded and monotonically increasing sequence, therefore they have a limit value. From a certain, sufficiently large n on, the $x_n, x_{n+1}, \ldots$ will differ from each other by at most ε, where ε is an arbitrarily small number:

$$|x_{n+1} - x_n| < \varepsilon, \quad \text{and } |cf(x_n)| < \varepsilon.$$

Since c was chosen to be different from zero, the limit of our sequence cannot be different from a root of $f(x)$; on the other hand, since the sequence can never grow beyond a and a is the next root to the right, the sequence $x_1, x_2, x_3, \ldots$ converges to this number.

In summary:

Theorem 1. If one takes a coefficient c whose magnitude is not greater than the reciprocal value of the largest gradient which the function $f(x)$ has in the interval in question, the sequence

$$x_1,\ x_2 = x_1 + cf(x_1),\ x_3 = x_2 + cf(x_2), \ldots$$

goes, depending on whether $cf(x_1)$ is positive or negative, to the next root to the right or left of x_1. The sequence is monotonically increasing or decreasing and goes to infinity only if there is no root to the right or left of x_1.

One sees that if $cf(x_1) > 0$, by means of the successively calculated $x_1, x_2, \ldots$ one approaches an existing root a from the left. By the way, the convergence will be the faster, the larger one has chosen c under consideration of (1), since x_{n+1} arises from x_n, by adding the $cf(x_n)$. It is also possible to change c from step to step, i.e. to choose a larger c when entering a region with a smaller slope (see the following example). When one has executed a few steps, one will calculate the value of $f(x_1')$ for a value x_1', which is slightly larger than the last calculated x_n, in order to limit the root also from the right. If this is negative, one can, if one wants to narrow the interval for the root still further, either calculate from the left still some steps, or starting from x_1' now similarly as in (2) form a decreasing sequence:

$$x_2' = x_1' + cf(x_1'),\ x_3' = x_2' + cf(x_2'), \ldots$$

which converges from the right towards a.

If $f(x)$ is negative for x_1, positive for x_1' (x_1 less than x_1'), then to get from x_1 to the next right root, do the analogous approach with a coefficient $-c < 0$:

$$x_2 = x_1 + cf(x_1),\ x_3 = x_2 + cf(x_2), \ldots$$

and to delimit the interval also from the other side:

$$x_2' = x_1' + cf(x_1'),\ x_3' = x_2' + cf(x_2'), \ldots$$

whereby one starts from an x_1', which lies a little to the right of the last calculated x_n'. and for which $f(x_1')$ is positive.

Here, simple roots are first considered, which are at the same time sign changes of the function $f(x)$. If one assumes that the equation has a multiple root, then one will probably examine as a rule the derivative of $f(x)$. However, the method presented here also allows to draw certain conclusions from the behavior of $f(x)$ alone. If the assumption is to be tested, whether e.g. between two numbers x_1 and x_1', where for example $f(x_1)$ and $f(x_1')$ are positive and $x_1 < x_1'$, there is a zero without sign change, then one forms the sequences :

$$\begin{aligned} x_2 &= x_1 + cf(x_1),\ x_3 = x_2 + cf(x_2), \ldots \\ x_2' &= x_1' + cf(x_1'),\ x_3' = x_2' + cf(x_2'), \ldots \end{aligned}$$

with the corresponding positive c (whereby again the c may be changed if necessary from step to step, but always (1) is to be considered). As long as these sequences approach each other without intersecting, a possible double root can only lie between them. If they intersect, there is certainly no double root.

Our method will now be explained by means of an example. The equation

$$\frac{1}{\chi} = \frac{4}{x^2}\frac{2(1-\cos x) - x\sin x}{x\cos x - \sin x}$$

gives for a certain class of cases the so called buckling condition for a clamped plane rectangular frame[2]. Here χ is a number which depends on the constants of the frame. For each χ there is a certain smallest x which determines the buckling load of the frame. We want to find this x value for $\chi = 1$ and thus consider the function

$$f(x) = \frac{4}{x^2}\frac{2(1-\cos x) - x\sin x}{x\cos x - \sin x} - 1.$$

[. . .]

2. Linear system.

The solution of linear equations with a larger number of unknowns forms a very difficult part of practical algebra. The solution by means of determinants is confusing and has still the special disadvantage that the accuracy of the result is already determined by the first step of the of the calculation. Since often small differences of large numbers (corresponding to small values of the determinants) arise in the course of the calculation, the accuracy can be impaired and this makes a repetition of the whole calculation necessary without using what has been achieved so far; indeed, it can happen that the most extensive calculating machine does not have enough digits to deliver a result of about three digits. - This procedure and the elimination procedure, which is similar to it, are competed by the procedures of the "approximation sequences" or iteration. They offer above all the advantage that one does not have to make a determination about the accuracy of the calculation from the outset, but in each stage of approximation, one can improve the accuracy if necessary always more and more. Moreover, errors do not propagate permanently, but are generally corrected automatically in the course of the calculation; at worst, they delay convergence.

We want to transfer the general iteration approach of §1 to linear equations. The system (1) to be solved, which we assume to be solvable (i.e. to have a determinant different from zero) is

$$L_1(x) \quad = \quad a_{1,1}x_1 + a_{1,2}x_2 + \cdots a_{1,n}x_n - r_1 = 0$$

[2]See R. v. Mises and J. Ratsersdorfer, this Zeitschr. vol. 6 (1926), p. 9.

$$\begin{aligned} L_2(x) &= a_{2,1}x_1 + a_{2,2}x_2 + \cdots a_{2,n}x_n - r_2 = 0 \\ \vdots &= \vdots \\ L_n(x) &= a_{n,1}x_1 + a_{n,2}x_2 + \cdots a_{n,n}x_n - r_n = 0 \end{aligned}$$

or written in vector form for short:

$$\mathfrak{A}\mathfrak{x} - \mathfrak{r} = 0 \qquad (1')$$

Here $\mathfrak{x}$ denotes the vector with the components $x_1, \ldots, x_n$; $\mathfrak{r}$ denotes the vector with components $r_1, \ldots, r_n$, and $\mathfrak{A}\mathfrak{x}$ represents the vector $\mathfrak{x}$ transformed with the matrix $\mathfrak{A}$, whose ith component is $a_{i,1}x_1 + a_{i,2}x_2 + \cdots a_{i,n}x_n$ $(i = 1, 2, \ldots, n)$. Start from any values $x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(2)}$ chosen arbitrarily or according to a conjecture about the approximate location of the solution and from this determine new quantities $x_1^{(2)}, x_2^{(2)}, \ldots, x_n^{(2)}$, according to the rule (2):

$$\begin{aligned} x_1^{(2)} &= x_1^{(1)} + c_1 L_1(x^{(1)}) \\ x_2^{(2)} &= x_2^{(1)} + c_2 L_2(x^{(1)}) \\ \vdots &= \vdots \\ x_n^{(2)} &= x_n^{(1)} + c_n L_n(x^{(1)}) \end{aligned}$$

One continues by using $x^{(2)}$ on the right-hand side instead of $x^{(1)}$ to determine the components of $x^{(3)}$ from them, etc. After $\nu - 1$ such "total steps", - where the calculation of n new quantities x^{ν} from n known quantities $x^{(\nu-1)}$ is called a "total step" - one has obtained a system $x^{(\nu)}$, from which one determines $x^{(\nu+1)}$ etc. The question is whether this process is a convergent one and whether the sequence of values $x^{(\nu)}$ converges to the solution of (1); more precisely, whether it is possible under certain conditions via the given system of equations to give a rule for the choice of c_i $(i = 1, 2, \ldots, n)$ - analogous to what was achieved in §1 in the case of one unknown - in such a way that the approximation sequence converges to the solution[3].

First, in generalization of the earlier thought process, a first sufficient convergence condition is as follows. If we denote by $x_1, \ldots, x_n$ a solution for which thus $L_i(x) = 0$ $(i = 1, 2, \ldots, n)$, and introduce quantities $z_i^{(\nu)}$ by

$$x_i^{(\nu)} - x_i = z_i^{(\nu)}$$

as the ith "error component" of the νth approximation, then from (2), by writing these equations for the upper index ν and $\nu + 1$ (instead of 1 and

[3]The process is well known. (See for example C. Runge "Praxis der Gleichungen", Leipzig 1921, p. 70 ff). However, a convergence condition is usually expressed as the quite indefinite requirement that the diagonal elements in the coefficient matrix should be "predominantly large".

2), we obtain a system (3) of equations homogeneous in the z,

$$\begin{aligned} z_1^{(\nu+1)} &= (1+c_1a_{1,1})z_1^{(\nu)} + c_1a_{1,2}z_2^{(\nu)} + \cdots + c_1a_{1,n}z_n^{(\nu)} \\ z_2^{(\nu+1)} &= c_2a_{2,1})z_1^{(\nu)} + (1+c_2a_{2,2}))z_2^{(\nu)} + \cdots + c_1a_{2,n}z_n^{(\nu)} \\ \vdots &= \vdots \\ z_n^{(\nu+1)} &= c_na_{n,1}z_1^{(\nu)} + c_na_{n,2}z_2^{(\nu)} + \cdots + (1+c_na_{n,n})z_n^{(\nu)} \end{aligned}$$

If the sums of the absolute values taken in columns are smaller than a real fraction μ (4):

$$\begin{aligned} |1+c_1\,a_{1,1}| &+ |c_2\,a_{2,1}| + \cdots + |c_n\,a_{n,1}| \le \mu < 1 \\ |c_1\,a_{1,2}| &+ |1+c_2\,a_{2,2}| + \cdots + |c_n\,a_{n,2}| \le \mu < 1 \\ \vdots &+ \vdots \\ |c_1\,a_{1,n}| &+ |c_2\,a_{2,n}| + \cdots + |1+c_n\,a_{n,n}| \le \mu < 1 \end{aligned}$$

it follows from (3) by adding the inequality (5):

$$\begin{aligned} |z_1^{(\nu+1)}| + |z_2^{(\nu+1)}| + \cdots + |z_n^{(\nu+1)}| &\le \mu[\,|z_1^{(\nu)}| + |z_2^{(\nu)}| + \cdots + |z_n^{(\nu)}|\,] \\ &\le \mu^{\mu}[\,|z_1^{(1)}| + |z_2^{(1)}| + \cdots + |z_n^{(1)}|\,] \end{aligned}$$

Thus it is shown that with each total step the sum of the absolute values of the error components becomes smaller and decreases after ν steps like the νth power of a real fraction, i.e. becomes arbitrarily small with indefinitely increasing ν, as soon as (4) is fulfilled. The same is true a fortiori for each individual error component[4].

The search for such coefficients c_i, for which (4) is fulfilled, can only be done in special cases, especially if in each row the diagonal entry $a_{i,i}$ is sufficiently predominant compared to the other entries $a_{i,j}$, $(i \neq j)$. In such a case, one will set

$$c_i = -\frac{1}{a_{i,i}}, \quad (i = 1, \ldots, n) \qquad (6)$$

to make the diagonal entries in the scheme (3) or (4) disappear.

The iteration (2) with the assumption (6) gets the form

$$x_i^{(\nu+1)} = -\frac{1}{a_{i,i}}[\,a_{i,1}x_1^{(\nu)} + a_{i,2}x_2^{(\nu)} + \cdots + a_{i,n}x_n^{(\nu)}\,] \qquad (7)$$

so that the ith equation is used to improve the ith unknown (solved for the ith unknown).

[4] A corresponding convergence condition for nonlinear systems of equations is given e.g. by C. Runge, Praxis der Gleichungen. Leipzig 1921, p. 69-70

[Note from the translator: This is not correct, it must be

$$x_i^{(\nu+1)} = -\frac{1}{a_{i,i}}[\, a_{i,1}x_1^{(\nu)} + \cdots + a_{i,i-1}x_{i-1}^{(\nu)} + a_{i,i+1}x_{i+1}^{(\nu)} + \cdots + a_{i,n}x_n^{(\nu)} - r_i \,] \qquad (7)$$

]

The condition (4) takes the form:

$$\sum_i{}' \left| \frac{a_{i,k}}{a_{i,i}} \right| \leq \mu < 1 \quad (k = 1, \ldots, n) \qquad (8)$$

where $\sum'$ means that the sum is only over $i \neq k$, i.e. only the column sums from the non-diagonal entries. In particular, of course, (8) is always satisfied if none of the $(n-1)$ terms in (8) exceeds the number $1/(n-1)$, (where the equal sign may not always apply), i.e., if in each row the diagonal entry is at least $(n-1)$ times as large as each of the other entries. Since any system of equations (1) with non-vanishing determinant can be ordered so that the diagonal entries have non-vanishing coefficients, and then by dividing each equation by the coefficient of the diagonal entry we obtain a system with all ones in the diagonal, the previous result is:

Theorem 2. If it is possible to determine coefficients $c_1, c_2, \ldots, c_n$ so that the n column sums (4) formed with them do not exceed a real fraction μ, so the iteration (2) formed with these c_i converges.

In particular, the iteration with $c_1, c_2, \ldots, c_n = -1$ always leads to the goal, if in the "fully divided" system with diagonal entries equal to one, the n sums of the amounts of the $(n-1)$ other entries are all less than one.

Before we proceed to give further convergence conditions, let us make a remark about the error estimate. For the practical calculation, the question of a useful error estimate at every stage of the calculation is more important - but usually even more difficult - than the question of convergence. An inexact but common method of error estimation in practical analysis can be applied to the problem of solving linear equations by iteration as follows. For two successive approximations $\mathfrak{x}^\nu$, $\mathfrak{x}^{\nu+1}$ and a solution $\mathfrak{x}$ according to the iteration approach (2), since $L_i(x) = 0$, it holds

$$x_i^{(\nu+1)} - x_i^{(\nu)} = c_i\,[\, L_i(x^{(\nu)}) - L_i(x) \,] = c_i\, \bar{L}_i(x^{(\nu)} - x) \qquad (9)$$

where $\bar{L}_i$, is obtained from L_i, by omitting the constant term. $\bar{L}_i(x)$ is thus also a term for the ith component $a_{i,1}x_1 + \cdots a_{i,n}x_n$ of the transformed vector $\mathfrak{A}\mathfrak{x}$. If then

$$x_i^{(\nu+1)} - x_i^{(\nu)} = \delta_i^{(\nu)} \quad (i = 1, \ldots, n) \qquad (10)$$

and if one sets again for $x_i^{(\nu)} - x_i = z_i^{(\nu)}$, (error of the νth approximation) one has:

$$\bar{L}_i(z^{(\nu)}) = \frac{\delta_i^{(\nu)}}{c_i}, \quad (i = 1, \ldots, n) \qquad (11)$$

Thinking of equation (11) solved for the unknown errors $z^{(\nu)}$, if A and $A_{i,\chi}$ denote the determinant and the adjoint subdeterminants from the $a_{i,\chi}$, respectively, we obtain

$$z_\chi^{(\nu)} = \sum_{i=1}^{n} \frac{\delta_i^{(\nu)}}{c_i} \frac{A_{i,\chi}}{A}, \quad |z_\chi^{(\nu)}| \leq \sum_{i=1}^{n} \left|\frac{\delta_i^{(\nu)}}{c_i}\right| \left|\frac{A_{i,\chi}}{A}\right| \leq \delta \sum_{i=1}^{n} \left|\frac{A_{i,\chi}}{A}\right| \qquad (12)$$

if δ is strictly larger than the largest of the $|\delta_i^{(\nu)}|/|c_i|$. If, after a few more steps, one has arrived at the point where the improvements brought by the ρth step, are on average only half as large as those of the νth step,

$$x_i^{(\rho+1)} - x_i^{(\rho)} = \delta_i^{(\rho)}, \quad |\delta_i^{(\rho)}| \sim \frac{1}{2}|\delta_i^{(\nu)}|$$

it follows that

$$|z_\chi^{(\nu)}| \leq \frac{1}{2}\delta \sum_{i=1}^{n} \left|\frac{A_{i,\chi}}{A}\right| \qquad (12')$$

If (12) and (12') were not inequalities but equations without an absolute value on the left, one could subtract (12') from (12) and get

$$z_\chi^{(\nu)} - z_\chi^{(\rho)} = x_\chi^{(\nu)} - x_\chi^{(\rho)} = \frac{1}{2}\delta \sum_{i=1}^{n} \left|\frac{A_{i,\chi}}{A}\right| = |z_\chi^{(\rho)}|.$$

The usual, unprovable assertion, at most made plausible by the above, is that the real error $z_\chi^{(\rho)}$ of the ρth approximation is of the same order of magnitude as the difference between the νth and ρth approximations, if the improvements brought about by the ρth step are on average half as great as those of the νth step.

An exact error approximation is possible if one assumes that condition (4) or (8) applies. Let the iteration be carried out so far that the improvement of the ith coordinate in the case of the νth iteration is such that:

$$|x_i^{(\nu+1)} - x_i^{(\nu)}| = |z_i^{(\nu)} - z_i^{(\nu+1)}| \leq \delta_i \qquad (14)$$

Then

$$\sum_{i=1}^{n} |z_i^{(\nu)} - z_i^{(\nu+1)}| \leq \delta_1 + \cdots + \delta_n = \sum_{i=1}^{n} \delta_i$$

and

$$\sum_{i=1}^{n} \{\, |z_i^{(\nu)}| - |z_i^{(\nu+1)}| \,\} = \sum_{i=1}^{n} |z_i^{(\nu)}| - \sum_{i=1}^{n} |z_i^{(\nu+1)}| \leq \sum_{i=1}^{n} \delta_i.$$

If one sets

$$\sum_{i=1}^{n} |z_i^{(\nu)}| = b, \quad \sum_{i=1}^{n} |z_i^{(\nu+1)}| = a, \quad \sum_{i=1}^{n} \delta_i = \delta,$$

then

$$b - a \leq \delta \text{ or } \frac{b}{a} \leq 1 + \frac{\delta}{a}.$$

On the other hand, (5) says that $a \leq \mu b$, so

$$\frac{1}{\mu} \leq \frac{b}{a} \leq 1 + \frac{\delta}{a}.$$

If one omits the middle term of the inequality, one obtains

$$1 + \frac{\delta}{a} \geq \frac{1}{\mu}, \quad a \leq \frac{\mu}{1-\mu}\delta$$

and

$$b = \sum_{i=1}^{n} |z_i^{(\nu)}|a + \delta \leq \frac{1}{1-\mu}\delta.$$

We thus obtain the following result which, again for simplicity, we express for a system with ones as diagonal coefficients.

Theorem 3. If the column sum of the absolute values of the non-diagonal entries in a system of equations with the diagonal coefficients equal to one is smaller than μ and one uses the usual iteration (7) (in which the ith equation is used to improve the ith unknown) such that that the difference of the νth approximation to the following one in the ith coordinate does not exceed δ_i, then the sum of the components of the real errors of the νth approximation is at most equal to

$$\frac{1}{1-\mu}\sum_{i=1}^{n}\delta_i.$$

It can be seen that the smaller is μ, the better is the bound. Our proposition can also be expressed in such a way that in the case under consideration the contributions of the individual improvements decrease substantially like the elements of a geometric sequence with the quotient μ.

As an example of this error estimation, consider the following system of equations[5]

$$\begin{aligned} 3x + 0.15y - 0.09z &= 6 \\ 0.08x + 4y - 0.16z &= 12 \\ 0.05x - 0.3y + 5z &= 20 \end{aligned}$$

The following approximations are calculated:

Here we can set $\mu = 0.11$; because the three absolute column sums of the "divided through" system are: $0.02 + 0.01 = 0.03$, $0.05 + 0.06 = 0.11$ and

[5]Runge-König, Vorlesungen über numerisches Rechnen [Lectures on Numerical Calculation]. Berlin 1924, p. 184.

	x	y	z
1. approximation	2	3	4
2. approximation	1.97	3.12	4.16
3. approximation	1.9688	3.127	4.1675
4. approximation	1.96868	3.12732	4.16793
5. approximation	1.96867	3.12734	4.16795

$0.03 + 0.04 = 0.07$; then $\frac{1}{1-\mu} = 1.13$. For $\delta_i^{(4)}$, that is for the difference of the fourth and fifth approximation, we find for the three variables 0.00001, 0.00002, 0.00002. The sum of these improvements is 0.00005 and our theorem shows that the sum of the real errors of the fourth approximation cannot exceed the number $0.00005 \cdot 1.13 = 0.00006$. For the third approximation the analogous consideration gives 0.00093 as an upper bound, while in fact the third decimal place in each coordinate is already correct at the third approximation.

On the other hand, the bound we have given may be reached under certain circumstances, as the following example shows, where μ cannot be set smaller than 0.5. If one goes with $c_1 = c_2 = -1$ in the equations

$$x + 0.5y = 2, \quad 0.5x + y = 2.5,$$

then the iteration formulas (7) yield

$$x^{(\nu+1)} = 2 - \frac{y^{(\nu)}}{2}, \quad y^{(\nu+1)} = 2.5 - \frac{x^{(\nu)}}{2}, \quad (\nu = 1, 2, \ldots).$$

If one sets $x^{(1)} = 0$, $y^{(1)} = 2.5$, then $x^{(2)} = 0.75$, $y^{(2)} = 2.5$. Then, $\delta_1 = 0.75$, $\delta_2 = 0$, $\delta_1 + \delta_2 = 0.75$,

$$\frac{\mu}{1-\mu}(\delta_1 + \delta_2) = 0.75, \quad \frac{\delta_1 + \delta_2}{1-\mu} = 1.50.$$

In fact, since $x = 1$, $y = 2$ is the exact solution, the absolute error sum of the first approximation is $1 + 0.5 = 1.5$; that of the second is $0.25 + 0.5 = 0.75$. Thus, the given bound is exactly reached, and it remains so in the following steps:

$$\begin{aligned} x^{(3)} &= 0.75,\ x^{(4)} = 0.9375,\ x^{(5)} = 0.9375,\ x^{(6)} = 0.984375,\ x^{(7)} = 0.984375, \\ y^{(3)} &= 2.125,\ y^{(4)} = 2.125,\ y^{(5)} = 2.03125,\ y^{(6)} = 2.03125,\ y^{(7)} = 2.0078125, \end{aligned}$$

etc.

We now turn to the discussion of other convergence cases. If the non-diagonal entries (in the system with ones on the diagonal) are generally of the order of $1/n$, while the columns have a sum that exceeds one, then

iteration (7) can still lead to the goal, if the inequality

$$\sum_{i,\chi}{}' \left(\frac{a_{i,\chi}}{a_{i,i}}\right)^2 < 1, \ (i \neq \chi) \qquad (16)$$

is satisfied, which we call Schmidt's condition because it is analogous to a convergence condition given by E. Schmidt for integral equations[6].

In vector form, the proof looks like this: Our system of equations is as above:

$$\mathfrak{A}\mathfrak{x} = \mathfrak{r} \quad \text{with } a_{i,i} = 1 \ (i = 1, \ldots, n)$$

Then the iteration with $c = -1$:

$$\mathfrak{x}^{(\nu+1)} = \mathfrak{x}^{(\nu)} - (\mathfrak{A}\mathfrak{x}^{(\nu)} - \mathfrak{r})$$

or, if again the "errors" $\mathfrak{z}^{(\nu)} = \mathfrak{x}^{(\nu)} - \mathfrak{x}$ are introduced:

$$\mathfrak{z}^{(\nu+1)} = \mathfrak{z}^{(\nu)} - \mathfrak{A}\mathfrak{z}^{(\nu)} = (\mathfrak{E} - \mathfrak{A})\mathfrak{z}^{(\nu)} = \mathfrak{K}\mathfrak{z}^{(\nu)} \qquad (17)$$

with $\mathfrak{E}$ the identity matrix and

$$\mathfrak{K} = \mathfrak{E} - \mathfrak{A}, \ k_{i,i} = 0, \ k_{i,\chi} = -a_{i,\chi} \ (i \neq \chi) \qquad (18)$$

So one has successively:

$$\mathfrak{z}^{(2)} = \mathfrak{K}\mathfrak{z}^{(1)}, \quad \mathfrak{z}^{(3)} = \mathfrak{K}\mathfrak{z}^{(3)} = \mathfrak{K}^2\mathfrak{z}^{(1)}, \ldots$$

where the elements of $\mathfrak{K}^2$, $\mathfrak{K}^3, \ldots$ are according to the multiplication rule for matrices:

$$k_{i,\chi}^{(2)} = \sum_{\rho} k_{i,\rho} k_{\rho,\chi}, \quad k_{i,\chi}^{(3)} = \sum_{\rho} k_{i,\rho}^{(2)} k_{\rho,\chi}, \ \ldots$$

It is to be shown that $k_{i,\rho}^{(\nu)}$ goes to zero with increasing ν.

To the equation

$$k_{i,\chi}^{(\nu)} = \sum_{\rho} k_{i,\rho}^{(\nu-1)} k_{\rho,\chi} = \sum_{\rho} k_{i,\rho}\, k_{\rho,\chi}^{(\nu-1)} \qquad (19)$$

one applies Schwarz's inequality:

$$|k_{i,\chi}^{(\nu)}|^2 < \sum_{\rho} [k_{i,\rho}^{(\nu-1)}]^2 \sum_{\rho} k_{\rho,\chi}^2 \qquad (20)$$

and from this by summing again over i and χ:

$$\sum_{i,\chi} |k_{i,\chi}^{(\nu)}|^2 < \sum_{i,\rho} [k_{i,\rho}^{(\nu-1)}]^2 \sum_{\chi,\rho} k_{\rho,\chi}^2 < \left[\sum_{\rho,\chi} k_{\rho,\chi}^2\right]^{\nu} = C^{\nu} \qquad (21)$$

[6]See Riemann-Weber, Die Differential- und Integralgleichungen der Mechanik und Physik [The Differential and Integral Equations of Mechanics and Physics]. 7th ed., vol. 1, Braunschweig 1925, p. 483.

Similarly

$$\begin{aligned} k_{i,\chi}^{(\nu+2)} &= \sum_{\rho} k_{i,\rho}^{(\nu+1)} k_{\rho,\chi} = \sum_{\rho} k_{\rho,\chi} \sum_{\lambda} k_{i,\lambda}^{(\nu)} k_{\lambda,\rho} = \sum_{\rho,\lambda} k_{\rho,\chi} k_{\rho,\lambda} k_{i,\lambda}^{(\nu)} \\ &= \sum_{\rho} k_{i,\rho} k_{\rho,\chi}^{(\nu+1)} = \sum_{\rho} k_{i,\rho} \sum_{\lambda} k_{\lambda,\chi} k_{\rho,\lambda}^{(\nu)} = \sum_{\rho,\lambda} k_{i,\rho} k_{\lambda,\chi} k_{\rho,\lambda}^{(\nu)}. \end{aligned}$$

Here, applying the same inequality again yields:

$$[k_{i,\chi}^{(\nu+2)}]^2 < \sum_{\rho,\lambda} [k_{i,\rho}\, k_{\lambda,\chi}]^2 \sum_{\rho,\lambda} [k_{\rho,\lambda}^{(\nu)}]^2 < C_1 C^{\nu} \qquad (22)$$

It can be seen that the iteration makes the quantities $k_{i,\chi}^{(\nu+2)}$, i.e. the vector of the error to converge towards zero, if the quantity C introduced here is smaller than one, i.e. if

$$\sum_{i,\chi} k_{i,\chi}^2 < 1.$$

This gives the above condition (16) for the coefficients $a_{i,\chi}$ of any linear system of equations. - This condition overlaps with the one expressed in Theorem 3. For example, for the matrix

$$\begin{pmatrix} 1 & 2/3 & 1/3 \\ 1/2 & 1 & 0 \\ 0 & 1/3 & 1 \end{pmatrix}$$

the convergence of the iteration according to Schmidt's condition is ensured, whereas Theorem 3 is not applicable. For the matrix

$$\begin{pmatrix} 1 & 3/4 \\ 3/4 & 1 \end{pmatrix}$$

the opposite is true. We have the following,

Theorem 4: The iteration in total steps with $c_i = -1/a_{i,i}$ according to equation (7) (or with $c_i = -1$ in the "divided through" system) converges, if for the coefficients $a_{i,\chi}$ of the given system of equations (1) the condition

$$\sum_{i,k} \frac{a_{i,k}^2}{a_{i,i}^2} < 1$$

is satisfied, where the sum is to be extended only over the combinations $i \neq k$.

Finally, the theoretically interesting case should be mentioned for which all coefficients c_i, are assumed to be equal from the beginning, i.e. the approach (29)

[Note from the translator: This must have been (23)]

$$\begin{aligned} x_1^{(\nu+1)} &= (1+c\,a_{1,1})x_1^{(\nu)} + c\,a_{1,2}x_2^{(\nu)} + \cdots + c\,a_{1,n}x_n^{(\nu)} - c\,r_1 \\ x_2^{(\nu+1)} &= c\,a_{2,1})x_1^{(\nu)} + (1+c\,a_{2,2}))x_2^{(\nu)} + \cdots + c\,a_{2,n}x_n^{(\nu)} - c\,r_2 \\ \vdots &= \vdots \\ x_n^{(\nu+1)} &= c\,a_{n,1}x_1^{(\nu)} + c\,a_{n,2}x_2^{(\nu)} + \cdots + (1+c\,a_{n,n})x_n^{(\nu)} - c\,r_n \end{aligned}$$

In vector form, if again the "errors" $\mathfrak{z}^{(\nu)}$ are introduced,

$$\mathfrak{x}^{(\nu+1)} = \mathfrak{x}^{(\nu)} + c(\mathfrak{A}\mathfrak{x}^{(\nu)} - \mathfrak{r}) = (\mathfrak{E} + c\mathfrak{A})\mathfrak{x}^{(\nu)} - c\mathfrak{r}, \quad \mathfrak{E} + c\mathfrak{A} = \mathfrak{K},\ \mathfrak{z}^{(\nu+1)} = \mathfrak{K}\mathfrak{z}^{(\nu+1)} \tag{29'}$$

The transformation which derives from the vector $\mathfrak{x}^{(\nu)}$ the new vector $\mathfrak{x}^{(\nu+1)}$ - from the error $\mathfrak{z}^{(\nu)}$ the new error $\mathfrak{z}^{(\nu+1)}$ - will lead the error vector to zero if and only if the eigenvalues of the matrix $\mathfrak{K} = \mathfrak{E} + c\mathfrak{A}$, which are the roots λ' of the equation:

$$\begin{vmatrix} 1+c\,a_{1,1}-\lambda' & c\,a_{1,2} & \cdots & c\,a_{1,n} \\ c\,a_{2,2} & 1+c\,a_{2,2}-\lambda' & \cdots & c\,a_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ c\,a_{n,1} & c\,a_{n,2} & \cdots & 1+c\,a_{n,n}-\lambda' \end{vmatrix} \tag{30}$$

all have absolute values smaller than one. If one writes, which is always possible because of $c \neq 0$,

$$1 + c\,a_{i,i} - \lambda' = c\left[a_{i,i} - \left(\frac{\lambda'}{c} - \frac{1}{c}\right)\right] = c(a_{i,i} - \lambda) \text{ with } \lambda' = \lambda\,c + 1,$$

then one sees that $\lambda' < 1$ means the same as $\lambda\,c + 1 < 1$ or

$$\left|\lambda + \frac{1}{c}\right| < \frac{1}{c} \tag{31}$$

But this means: the eigenvalues λ of the original matrix $\mathfrak{A}$, i.e. the roots of the equation

$$\begin{vmatrix} a_{1,1}-\lambda & a_{1,2} & \cdots & a_{1,n} \\ a_{2,2} & a_{2,1}-\lambda & \cdots & c\,a_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ c\,a_{n,1} & c\,a_{n,2} & \cdots & a_{n,n}-\lambda \end{vmatrix} \tag{32}$$

must satisfy the relation given by (31) with the coefficient c.

When all roots of (32) are on the left or all on the right side of the imaginary axis and one chooses c according to the sign positive in that, negative in this case and $|1/c|$ is equal to the radius of a circle which contains all eigenvalues and touches the imaginary axis at the origin, then (31) is

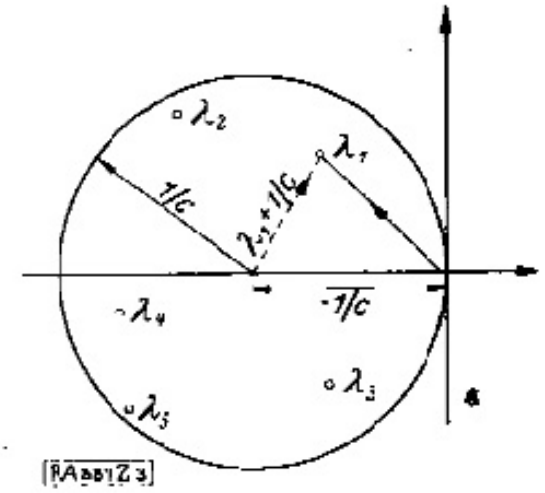

Fig. 3

always satisfied (cf. Fig. 3; this is drawn for the case that all λ have negative real part and c is therefore chosen positive). Since $1/c$ must grow with ρ, it can be seen that the convergence of the method becomes slower, the larger the smallest circle is. We have

Theorem 5: The iteration procedure (29) with equal coefficients c always converges if all eigenvalues of the matrix of the system lie on one side of the imaginary axis, provided that c is chosen as in Fig. 3. $-1/c$ is the (real) center point abscissa of a circle which touches the imaginary axis at the origin and contains all the eigenvalues.

If the matrix of $a_{i,\chi}$ is positive - or negative - definite, the mentioned condition is always fulfilled, i.e. a suitable c can be found. Whether it is practically possible to determine this c numerically is a question which cannot be generally answered in the affirmative. As a rule, in cases where neither Theorem 2 nor Theorem 4 is applicable, it will be advisable to use the methods discussed in the following section.

3. Iteration in single steps.

In an important class of cases we can a somewhat different iteration procedure can be carried out, which we call "iteration in single steps" in contrast to the iteration in "total steps" considered so far. We arrive at a unified conception of both types of iteration, by first generalizing the procedure (2) of §2. We want to assume that the iteration coefficients c_i, (as it was foreseen in §1 for the case of one variable) also depend on the index ν of the step:

$$x_i^{(\nu+1)} = x_i^{(\nu)} + c_i^{(\nu)} L_i(x^{(\nu)}) \qquad (1)$$

If one sets here $c_i^{(1)} = c_i^{(2)} = \cdots = c_i^{(n)}$, then one has the earlier procedure of the iteration in "total steps". We now use another special approach: at the first step $\nu = 1$ let all $c_i^{(1)}$ except $c_1^{(1)}$ be zero, so that only $x_1^{(1)}$ is changed, at the second step let all $c_i^{(2)}$ $(i \neq 2)$ equal zero, at the third all $c_i^{(3)}$ $(i \neq 3)$ and so on. In the $(n+1)$th step, the value of $c_1^{(n+1)}$ is again taken equal to $c_1^{(1)}$, and all $c_i^{(n+1)}, (i \neq 1)$ equal to zero, etc. So the value of x_i is improved only at the $(n+i)$th, $(2n+i)$th and so on. To simplify the notation we put:

$$\begin{aligned} c_1^{(1)} &= c_1^{(n+1)} = c_1^{(2n+1)} = \cdots = c_1 \\ c_2^{(2)} &= c_2^{(n+2)} = c_2^{(2n+2)} = \cdots = c_2 \\ \vdots &= \vdots \\ c_n^{(n)} &= c_n^{(2n)} = c_n^{(3n)} = \cdots = c_n \end{aligned}$$

and write $x_i^{(\nu)}$ for short for the quantity occurring n times, which according to the original notation should carry the upper indices $(\nu-1)n+i$, $(\nu-1)n+i+1$, $(\nu-1)n+i+2$, …, $\nu n+i-1$. The iteration is then described by the following formula (2) which can be understood without reference to (1):

$$\begin{aligned} x_1^{(2)} &= x_1^{(1)} + c_1\, L_1(x_1^{(1)}, x_2^{(1)}, \ldots, x_{n-1}^{(1)}, x_n^{(1)}) \\ x_2^{(2)} &= x_1^{(1)} + c_2\, L_2(x_1^{(2)}, x_2^{(1)}, \ldots, x_{n-1}^{(1)}, x_n^{(1)}) \\ \vdots &= \vdots \\ x_n^{(2)} &= x_n^{(1)} + c_n\, L_n(x_1^{(2)}, x_2^{(2)}, \ldots, x_{n-1}^{(2)}, x_n^{(1)}) \end{aligned}$$

In the same way, the new $x_i^{(3)}$ are derived from the $x_i^{(2)}$ by inserting the $x_i^{(2)}$ on the right, and so on. The difference with the iteration used in the previous section is that when calculating the $(\nu+1)$th approximations $x_i^{(\nu+1)}$ $(i = 1, \ldots, n)$, the values $x_i^{(\nu)}$ are not used in all n equations, but in the χth equation on the right already the "newest" values determined immediately before by the preceding $(\chi-1)$ equations are used, thus $x_1^{(\nu+1)}$, $x_2^{(\nu+1)}$, …, $x_{\chi-1}^{(\nu+1)}$ serve for the calculation of $x_\chi^{(\nu+1)}$.

We want to show that in a practically very important class of cases this iteration method converges if one sets $c_i = -1/a_{i,i}$, where $a_{i,\chi}$, are, as always, the equation coefficients[7]. The iteration approach for this determination is as follows (3), since the first term on the right in (2) is canceled out against a later one:

$$x_1^{(\nu+1)} = -\frac{1}{a_{1,1}}[a_{1,2}x_2^{(\nu)} + a_{1,3}x_3^{(\nu)} + \cdots + a_{1,n-1}x_{n-1}^{(\nu)} + a_{1,n}x_n^{(\nu)} - r_1]$$

[7]This method of "iteration in single steps" was developed by Ph. L. Seidel Münch. Ak. Abb. 1874, 3. Abh. p. 81 to 108.

$$
\begin{aligned}
x_2^{(\nu+1)} &= -\frac{1}{a_{2,2}}[a_{2,1}x_1^{(\nu+1)} + a_{2,3}x_3^{(\nu)} + \cdots + a_{2,n-1}x_{n-1}^{(\nu)} + a_{2,n}x_n^{(\nu)} - r_2] \\
\vdots &= \vdots \\
x_n^{(\nu+1)} &= -\frac{1}{a_{n,n}}[a_{n,1}x_1^{(\nu+1)} + a_{n,3}x_3^{(\nu+1)} + \cdots + a_{n,n-1}x_{n-1}^{(\nu+1)} + a_{n,n}x_n^{(\nu)} - r_n]
\end{aligned}
$$

The calculation of the approximations must be done in the order given here, i.e. the $(\nu+1)$th approximation of the χth unknown can only be determined when the $(\nu+1)$th approximation for the 1st, 2nd,..., $(\chi-1)$th unknown has already been calculated.

This approach can be interpreted in such a way that the ith equation of the given system is solved for the ith unknown x_i, whereby one sets the $(\nu+1)$th values $x_1^{(\nu+1)}$, $x_2^{(\nu+1)}$, ..., $x_{i-1}^{(\nu+1)}$ for the first $(i-1)$ unknowns and the νth values $x_{i+1}^{(\nu)}$, $x_{i+2}^{(\nu)}$, ..., $x_n^{(\nu)}$ for the further ones. The iteration thus explained converges, as we shall show, whenever the equation coefficients $a_{i,\chi}$, are symmetric, $a_{i,\chi} = a_{\chi,i}$, and whenever the quadratic form $2Q = \sum_{i,\chi} a_{i,\chi}\, x_i\, x_\chi$ has always positive values when the variables take any nonzero values. Such a form is called positive definite. The coefficient matrix of such a form has, as it is well known, a positive determinant (which ensures the unambiguous solvability of (1)). Furthermore, the diagonal coefficients $a_{i,i}$ are certainly positive, (For if $a_{1,1} < 0$, for example, the term $a_{1,1}x_1^2$, which, if x_1 were sufficiently increased, outweighs all other terms, would make the value of Q negative; but if $a_{1,1} = 0$, $Q = 0$ for $x_2 = x_2 = \cdots = x_n = 0$ and any value of x_1). The following theorem is to be proved:

Theorem 6 If the coefficients $a_{i,\chi}$ of the given linear system of equations are symmetric, and if the quadratic form formed with them is positive definite, the iteration in single steps, where $c_i = -1/a_{i,i}$ is set, i.e. where only the ith equation, according to (3), is used each time to improve the ith unknown, always converges.

For the proof[8], we consider the following function in $x_1, \ldots, x_n$

$$
F(x_1, \ldots, x_n) = \frac{1}{2}\sum_{i,\chi} a_{i,\chi}\, x_i\, x_\chi - \sum_{i=1}^{n} r_i\, x_i = Q - \sum_{i=1}^{n} r_i\, x_i \qquad (4)
$$

The crucial point is that, as one can see immediately, the given linear equations are identical with

$$
\frac{\partial F}{\partial x_1} = 0,\ \frac{\partial F}{\partial x_2} = 0, \ldots, \frac{\partial F}{\partial x_n} = 0 \qquad (5)
$$

i.e., for a system of x-values satisfying our equations, F assumes a stationary value (maximum, minimum or saddle value). It can be further shown that

[8] A proof that agrees with the following was sent to the editors some time ago by Mr. E. Trefftz; its publication was only omitted in consideration of the report in preparation.

this value must be a minimum. Because if one gives sufficiently large values to the magnitudes of the x_i (or also only one of them), then the quadratic component Q of F outweighs the linear one and this is increasing with increasing x_i, since, according to the assumption, the quadratic form is positive definite, positive for all x_i different from zero and proportional to the squares of the x_i. Therefore, the function Foutside a certain range B grows more and more with increasing contributions of the x_i; thus, a place of stationary value can only be in the interior of B. As a continuous function bounded in B, F must have a minimum value which certainly does not lie at the boundary, and since our linear equations have only one solution system, the solution of our system must yield the location of this minimum.

It is thus shown that the function F certainly has one minimum, and only one, under the condition that its quadratic component Q cannot take negative values. As the train of thought of the proof shows, the existence of a minimum is only ensured if F itself is always positive for all values of the independent variables (except $x_i = 0$). This remark will be used in the following section. (By the way, it can be easily seen that if F has the property of being positive definite, Q must have it too).

In any case, the solution of equations (5) is equivalent to finding the minimum point of F. The convergence of our iteration (3) is a monotonic one in that it can be shown that each step taken according to (3) decreases the value of F. For the unknowns $x_1, \ldots, x_{i-1}$ the $(\nu+1)$th approximations are known, for $x_{i+1}, \ldots, x_n$ the νth approximations are known, and now the step follows, which replaces $x_i^{(\nu)}$ by $x_i^{(\nu+1)}$. The new value of F differs then from the old one only in the members which contain x_i. The increase is according to (4):

$$\begin{aligned}\Delta F = \frac{1}{2}a_{i,i}[(x_i^{(\nu+1)})^2 \quad &- \quad (x_i^{(\nu)})^2] + \{\sum_{\chi=1}^{i-1} a_{\imath,\chi}\, x_\chi^{(\nu+1)} \\ &+ \quad \sum_{\chi=i+1}^{n} a_{\imath,\chi}\, x_\chi^{(\nu)} - r_i\}\,[x_i^{(\nu+1)} - x_i^{(\nu)}].\end{aligned}$$

Now, from (3), for the expression in the curly brace, just put $-a_{i,i}\, x_i^{(\nu+1)}$; so, it becomes

$$\begin{aligned}\Delta F &= (x_i^{(\nu+1)} - x_i^{(\nu)})\left[\frac{1}{2}a_{i,i}(x_i^{(\nu+1)} + x_i^{(\nu)}) - a_{i,i}\, x_i^{(\nu+1)}\right] \\ &= -\frac{1}{2}a_{i,i}[x_i^{(\nu+1)} - x_i^{(\nu)}]^2 < 0 \qquad (6)\end{aligned}$$

since, as it was proved earlier, $a_{i,i}$ is positive. Thus, at each step of iteration (3), the expression F given by (4) is reduced. Since F has a finite minimum, the reduction cannot continue indefinitely, i.e. ΔF must converges towards 0. According to (6) it follows that the absolute value $|x_i^{(\nu+1)} - x_i^{(\nu)}|$ must approach 0 for each i, i.e. the procedure converges.

As an example, we consider a system of six inhomogeneous equations which have emerged from a minimum problem[9] - of order $n = 6$ - whose matrix is therefore certainly positive definite:

$$\begin{aligned}
651.8x - 239.2y + 94.6z - 188u + 148v + 58w &= 431.5,\\
-119.6x + 8867y - 961z - 226u - 515v + 186w &= 188.5,\\
94.6x - 1922y + 24390z + 474u - 4820v + 592w &= 82.7,\\
-93.9x - 226y + 237z + 48100u - 23370v - 2015w &= 120.0,\\
74.2x - 515y - 2410z - 2370u + 78600v - 6420w &= 52.5,\\
58x + 372y + 592z - 4030u - 12840v + 158500w &= 33,5.
\end{aligned}$$

From this system, a symmetric one is obtained when $x/2$, $z/2$, $w/2$ are considered as unknowns in place of x, z, w.

[Note from the translator: This is not true because of the entries $a_{1,4}$ and $a_{4,1}$, but the matrix is strictly diagonally dominant.]

When solving according to our iteration method, we first set values for the first approximate values x, y, z of the first three unknowns, which we obtained by applying the same minimum problem with the approximation degree $n = 3$. This corresponds to a system of three equations with the unknowns x, y, z, whose coefficient matrix is in the upper left corner of the six-row matrix and whose right-hand sides match the first three numbers of the above right-hand side of the six-part system. For this three-part system, the following table arises:

[. . .]

The last values found are used as first approximations of the first three unknowns in the six-part system and obtained following the iteration:

[. . .]

The geometric meaning of the specified method is as follows: In the $(n+1)$-dimensional space, in which F forms a "paraboloid" as a function of x_1 to x_n, one goes to the first step, x_1, improved by the first equation the paraboloid, within the two-dimensional hyperplane, which is defined by the $(n-1)$ equations $x_2 = \text{const}, \ldots, x_n = \text{const}$ is determined up to the lowest point. After all, the first equation states that F is made a minimum when x_2 to x_n, are held fixed. The x_1, of the found point gives the improved value of the first unknown, Then one goes, by changing only x_2, in a two-dimensional plane, which contains the just found point, in the same way further, and so on.

It makes sense to proceed occasionally in such a way that one chooses two-dimensional, multi-dimensional hyperplanes which are fixed by setting less than $(n-1)$ coordinates as constants. In any case, one will always reach

[9]The approach is taken from a paper by W. Ritz, Crelle's Journal 185 (1909), pp. 1 to 61.

deeper points lying in the paraboloid, if one chooses the free coordinates in such a way that they satisfy the equations assigned to them. The procedure is then as follows: One chooses values for the third to nth unknown values $x_3^{(1)}, x_4^{(1)}, \ldots, x_n^{(1)}$, and determines the values $x_1^{(2)}, x_2^{(2)}$ by solving the first and second equation simultaneously according to the first and second coordinate, then one takes $x_1^{(2)}, x_2^{(2)}$, the former values of the sixth to nth unknowns $x_6^{(1)}, \ldots, x_n^{(1)}$ and, entering these $(n-3)$ quantities into the 3rd, 4th and 5th equations, determines from these three equations values $x_3^{(2)}, x_4^{(2)}, x_5^{(5)}$ of the third to fifth variables, and so on. We call this procedure, which is somewhat different from the original iteration, "iteration in groups" and formulate:

Theorem 7. A linear system of equations with a positive definite coefficient matrix can be solved by "iteration in groups" using to improve the unknowns $x_\alpha, x_\beta, x_\gamma, \ldots$ the equations numbered $\alpha, \beta, \gamma, \ldots$ as a simultaneous system of equations and with $\alpha, \beta, \gamma, \ldots$ in a cyclic sequence through all equation numbers.

This procedure will be used practically only if any groups of two, three, four equations appear to be particularly suitable for simultaneous solution by the arithmetic constitution of the system.

The importance of theorems 6 and 7 lies in the fact that a large part of the practically occurring systems of equations, especially those of statics, possess the property stated here. In addition, it can be shown that any linear system of equations which can be solved at all can be transformed in such a way that its coefficients satisfy the conditions of the theorems[10]. Namely, one obtains a system equivalent to the given one, by performing a simple transformation known from the balancing calculus, but which can be explained without any extraneous element by the following prescription: Multiply the first of the given n equations by $a_{1,1}$, the second by $a_{2,1}$,... the last by $a_{n,1}$, and add the n equations thus multiplied. The result is a first new equation - naturally dependent on the given equations. A second new equation is obtained by multiplying the χth of the original equations by $a_{\chi,2}$, $(\chi = 1, \ldots, n)$, and then adding these n equations, and so on. Thus one obtains n new equations, which are linearly independent of each other, if the original equations were so. As coefficients of these new equations are the sums introduced by Gauss in his treatise on the method of least squares, which are also called Gaussian bracket expressions:

$$[1\,1] = \sum_{\chi=1}^{n} a_{\chi,1}^2, \quad [1\,2] = [2\,1] = \sum_{\chi=1}^{n} a_{\chi,1}\, a_{\chi,2}, \ \ldots$$

Generally,

$$[i\,i] = \sum_{\chi=1}^{n} a_{\chi,i}\, a_{\chi,j} \ \text{ and } [i;\, r] = \sum_{\chi=1}^{n} a_{\chi,i}\, r_\chi, \quad (i, j = 1, 2, \ldots, n) \qquad (7)$$

[10]This idea, too, can already be found in the cited treatise by Ph. L. Seidel.

With the help of the calculating machine the expressions, which consist only of sums of products, can be formed without difficulty. However, $\frac{n(n+1)}{2} + n = \frac{n-n+3=}{2}$ of such brackets are to be calculated. The new equations (8) equivalent to the given ones are:

$$\begin{aligned} 0 &= [1\,1]x_1 + [1\,2]x_2 + \cdots + [1\,n]x_n - [1;\,r] \\ 0 &= [2\,1]x_1 + [2\,2]x_2 + \cdots + [2\,n]x_n - [2;\,r] \\ \vdots &\quad\; \vdots \\ 0 &= [n\,1]x_1 + [n\,2]x_2 + \cdots + [n\,n]x_n - [n;\,r] \end{aligned}$$

From (7) we see that the coefficients of (8) are symmetric. That they are also positive definite can be seen by the following consideration: The quadratic function, which can never become negative as a sum of squares,

$$Q = \left(\sum_{\chi=1}^{n} a_{1,\chi}\, x_\chi\right)^2 + \left(\sum_{\chi=1}^{n} a_{2,\chi}\, x_\chi\right)^2 + \cdots + \left(\sum_{\chi=1}^{n} a_{n,\chi}\, x_\chi\right)^2 \qquad (9)$$

is nothing else than the quadratic form belonging to (8). For example, it is the coefficient of x_1^2 in (9) equal to $a_{1,1}^2 + a_{2,1}^2 + \cdots + a_{n,1}^2$ - one member comes from each of the n brackets in (9) -, and this is $[1\,1]$; the coefficient of $x_1\,x_2$, likewise becomes equal to $2[1\,2]$, etc. Thus one has the general theorem:

Theorem 8: As soon as the conditions of Theorems 6 and 7 are not fulfilled for a system of equations, it is always possible to replace it by an equivalent system for which these conditions are fulfilled by forming the Gaussian bracket expressions (7) from the coefficients of the original system and using them to form the system (8).

4. Application in statics.

[. . .]

5. Homogeneous system of linear equations. Smallest eigenvalue.

[. . .]

6. Largest eigenvalues and eigensolutions.

[. . .]