# NATIONAL BUREAU OF STANDARDS REPORT

2076

## SOLVING LINEAR ALGEBRAIC EQUATIONS CAN BE INTERESTING

by

George E. Forsythe


U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

U. S. DEPARTMENT OF COMMERCE
Charles Sawyer, *Secretary*

NATIONAL BUREAU OF STANDARDS
A. V. Astin, *Director*

# THE NATIONAL BUREAU OF STANDARDS

The scope of activities of the National Bureau of Standards is suggested in the following listing of the divisions and sections engaged in technical work. In general, each section is engaged in specialized research, development, and engineering in the field indicated by its title. A brief description of the activities, and of the resultant reports and publications, appears on the inside of the back cover of this report.

**Electricity.** Resistance Measurements. Inductance and Capacitance. Electrical Instruments. Magnetic Measurements. Applied Electricity. Electrochemistry.

**Optics and Metrology.** Photometry and Colorimetry. Optical Instruments. Photographic Technology. Length. Gage.

**Heat and Power.** Temperature Measurements. Thermodynamics. Cryogenics. Engines and Lubrication. Engine Fuels. Cryogenic Engineering.

**Atomic and Radiation Physics.** Spectroscopy. Radiometry. Mass Spectrometry. Solid State Physics. Electron Physics. Atomic Physics. Neutron Measurements. Infrared Spectroscopy. Nuclear Physics. Radioactivity. X-Rays. Betatron. Nucleonic Instrumentation. Radiological Equipment. Atomic Energy Commission Instruments Branch.

**Chemistry.** Organic Coatings. Surface Chemistry. Organic Chemistry. Analytical Chemistry. Inorganic Chemistry. Electrodeposition. Gas Chemistry. Physical Chemistry. Thermochemistry. Spectrochemistry. Pure Substances.

**Mechanics.** Sound. Mechanical Instruments. Aerodynamics. Engineering Mechanics. Hydraulics. Mass. Capacity, Density, and Fluid Meters.

**Organic and Fibrous Materials.** Rubber. Textiles. Paper. Leather. Testing and Specifications. Polymer Structure. Organic Plastics. Dental Research.

**Metallurgy.** Thermal Metallurgy. Chemical Metallurgy. Mechanical Metallurgy. Corrosion.

**Mineral Products.** Porcelain and Pottery. Glass. Refractories. Enameled Metals. Concreting Materials. Constitution and Microstructure. Chemistry of Mineral Products.

**Building Technology.** Structural Engineering. Fire Protection. Heating and Air Conditioning. Floor, Roof, and Wall Coverings. Codes and Specifications.

**Applied Mathematics.** Numerical Analysis. Computation. Statistical Engineering. Machine Development.

**Electronics.** Engineering Electronics. Electron Tubes. Electronic Computers. Electronic Instrumentation.

**Radio Propagation.** Upper Atmosphere Research. Ionospheric Research. Regular Propagation Services. Frequency Utilization Research. Tropospheric Propagation Research. High Frequency Standards. Microwave Standards.

**Ordnance Development.**
**Electromechanical Ordnance.**
**Electronic Ordnance.**
These three divisions are engaged in a broad program of research and development in advanced ordnance. Activities include basic and applied research, engineering, pilot production, field testing, and evaluation of a wide variety of ordnance matériel. Special skills and facilities of other NBS divisions also contribute to this program. The activity is sponsored by the Department of Defense.

**Missile Development.** Missile research and development: engineering, dynamics, intelligence, instrumentation, evaluation. Combustion in jet engines. These activities are sponsored by the Department of Defense.

- Office of Basic Instrumentation
- Office of Weights and Measures.

# NATIONAL BUREAU OF STANDARDS REPORT

**NBS PROJECT** **NBS REPORT**

1101-10-5100 November 18, 1952 2076

SOLVING LINEAR ALGEBRAIC EQUATIONS
CAN BE INTERESTING

by

George E. Forsythe[1]

National Bureau of Standards

PREPRINT

[1]Sponsored in part by the Office of Naval Research, USN.

CONTENTS

# SOLVING LINEAR ALGEBRAIC EQUATIONS CAN BE INTERESTING

George E. Forsythe[1]

1. Introduction. The subject of this talk is mathematically a lowly one. Consider a system of n linear algebraic equations in n unknowns, written

$$Ax = b \quad . \tag{1}$$

Here A is a square matrix of order n, whose elements are given real numbers $a_{ij}$ with a determinant $d(A) \neq 0$; x and b denote column vectors, and the components of b are given real numbers. (Complex numbers would offer no essential difficulty.) It is desired to calculate the components of the unique solution $x = A^{-1}b$; here $A^{-1}$ is the inverse of A.

Such problems arise in the most diverse branches of science and technology, either directly (e.g., the normal equations of the least-squares adjustment of observations) or in an approximation to another

---

A talk delivered (under a similar title) before the Eugene meeting of the American Mathematical Society, June 21, 1952, by invitation of the Committee to Select Hour Speakers for Western Sectional Meetings; received by the editors , 1952.

[1]Sponsored in part by the Office of Naval Research, USN.

problem (e.g., the difference-equation approximation to a self-adjoint boundary-value problem for a partial differential equation). These two are very frequent sources of numerical systems; note that $A > 0$ (i.e., A is symmetric and positive definite) in both examples. The order n is considered to range from perhaps 6 or 8 up to as large a number as can be handled.

Stearn [111], for instance, mentions the solution of a system of order 2300 by the U.S. Coast and Geodetic Survey. The accuracy demanded of an approximate solution $\xi$ varies widely; even the function which is to measure the accuracy of $\xi$ varies or is unknown. Some "customers" want to make the length $|b - A\xi|$ small; some, $|\xi - A^{-1}b|$; others have apparently thought only in terms of getting $A^{-1}b$ exactly.

We all know that each component of the solution $A^{-1}b$ can be expressed as a quotient of determinants by Cramer's rule. We have all evaluated determinants of orders 3, 4, and possibly 5, with $a_{ij}$ integers; it is quite easy and rather boring. I therefore suspect that the average mathematician damns the practical solution of (1) as being both trivial and dull.

One defense of the numerical analyst (preface of [81]) is to show that in many practical cases (say for decimal fractions $a_{ij}$ and $n \geq 10$) getting $A^{-1}b$ efficiently (or at all) is actually not trivial, but demands both know-how and planning. This point is especially well taken against those whose program for solving (1) would be to evaluate $n + 1$ determinants from their definition, employing $n!$ multiplications per determinant. For $n = 26$, for example,

$(n + 1)! \doteq 10^{28}$, a number of multiplications which would take the SWAC[2] some $10^{17}$ years. Actually, only about $(1/3)n^3$ multiplications

[2] National Bureau of Standards Western Automatic Computer, an electronic machine easily capable of 2000 multiplications per second. See [61] for a description, now out of date.

are needed to solve (1); see Bodewig [12]. For $n = 26$, $(1/3)n^3 \doteq 6000$, and the multiplications would take the SWAC about 3 seconds.

It is my hope, on the other hand, to arouse the mathematician's interest by showing (section 2) the diversity of approaches to the solution of (1), and by mentioning (sections 3 to 6) some problems associated with selected iterative methods. The newest process on the roster, the method of conjugate gradients, is outlined in section 7. Section 8 touches on the difficult general subject of errors and "condition," while a few words are hazarded in section 9 about the effect of machines on methods.

Whether or not the subject proves interesting, the bibliography is intended to make the field look at least respectable! It is a representative condensation of the 500-odd titles in the author's file, most of which are in [27]. There are bibliographies on related subjects in Collatz [16], Dwyer [20], Engineering Research Associates [22], Frame [36], Frankel [37], Franklin and Hankam [38a], Harvard Computation Laboratory [47], Higgins [55], Kuroš, Markuševič, and Raševskiĭ [71], Motzkin [82], Schwerdtfeger [106], and Taussky [116].

It is remarkable how little is really understood about most of the methods for solving (1). They are being used nevertheless, and yield answers. This disparity between theory and practice appears to be typical of the gulf between the science and the art of numerical analysis.

2. Survey of methods. It usually surprises the uninitiated to learn the variety of methods actually used to solve systems (1). Surveys of methods are given by Bodewig [12], Dwyer [20], Faddeeva [23], Forsythe [26], Fox [33, 34], Frazer, Duncan and Collar [39], Hotelling [58], Householder [59, 60], Zurmühl [131], and others, but the spectrum of known methods is considerably broader than any of these surveys suggests. A classification of elimination methods is given by Jensen [65]. A tentative classification of all known methods, with a bibliography of about 450 titles, is in [27].

As mentioned in section 1, one can solve a system (1) explicitly by use of determinants, while explicit solutions in other forms are sometimes available. Friedrich and Jenne [40], for example, describe the use of continued fractions.

The best known methods are based on systematic elimination of unknowns from equations in the system (1) in the fashion of high-school algebra, as described by Gauss [42]. The elimination amounts to triangularizing the matrix A by premultiplying it by a triangular matrix, as Banachiewicz [5] and Turing [122] point out. The process can be rendered very efficient numerically by consolidating operations; see, for example, Benoit [7], Dwyer [20], and Turing [122].

When A is positive definite, the method is equivalent to the successive orthogonalization of the unit vectors in the A metric by the Gram-Schmidt process [99]. With other metrics the orthogonalization process yields different methods for solving (1). All these elimination methods can also be performed on any matrix of submatrix blocks formed by partitioning A, an idea suggested by Boltz [13] and based on relations given by Schur [105]. The various elimination methods are direct, a term defined at the end of section 2.

There is a group of direct methods related to the characteristic polynomial $\phi$ of some matrix. For example, if one can learn that $\phi(A) \equiv c_n A^n + \cdots + c_1 A + I \equiv 0$, then, as Bingham [8] notes, one can compute $A^{-1}b = -c_n A^{n-1}b - \cdots - c_2 Ab - c_1 b$. Similar remarks apply when $\phi(H)$ is known for an operator H associated with the solution of (1); see section 4. There are related methods involving the successive orthogonalization of the vectors $Ax_o, A^2x_o, \cdots, A^nx_o$ in the I, A, $A^{-1}$, or other metrics; one of these is given in section 7.

There is an unpublished direct method of Hans Lewy using the theory of congruences in n dimensions, applicable mainly when the components of $A^{-1}b$ are integers. It is based on the use of stencils for solving the system $A^*x = b^*$, where $a^*_{ij} = 0$ (if $a_{ij}$ is even) and $a^*_{ij} = 1$ (if $a_{ij}$ is odd), and where $b^*_i$ is similarly defined.

Of a quite different nature is a group of iterative processes devised by Jacobi [63], Nekrasov [88], Richardson [98], and others, and subsumed by Cesari [15], Wittmeyer [127], and Geiringer [44] in a general class of linear iterative processes to be discussed in

section 3. In these methods one starts with an arbitrary first vector $x_0$ approximating $A^{-1}b$. For $k = 0, 1, 2, \cdots$, the components of the k-th approximant vector $x_k$ are systematically corrected, often in the cyclic order $1, 2, \cdots, n$, but sometimes in blocks or otherwise. After one cycle of the corrections has been completed, the components of the resulting vector $x_{k+1}$ will have been obtained by solving a matrix equation $Bx_{k+1} + Cx_k = b$ for $x_{k+1}$, where $B + C = A$. The different linear processes are distinguished by the choice of B, which must be an easily invertible matrix. For example, B can be the lower triangle of A [88], the diagonal of A [63], a scalar matrix [98], diagonal blocks of A (von Mises and Pollazek-Geiringer [123], Hertwig [49]), etc. (When B is the lower triangle of A, the iterative method is commonly called the 'Seidel process,' or the 'Gauss-Seidel process.' But, as Ostrowski [91] points out, Seidel [107] mentions the process but advocates not using it. Gauss nowhere mentions it. Nekrasov [88] studies its properties and says it is Seidel's process. It will henceforth be called the 'cyclic single-step process.") Under appropriate conditions mentioned in section 3, $x_k$ approaches $A^{-1}b$, but ordinarily no $x_k$ equals $A^{-1}b$.

Another group of iterative processes includes those of Gauss [41], Seidel [107], Southwell [109, 110], Motzkin [2], and others. These are called "relaxation methods" in [109, 110]. They are discussed by Black and Southwell [10], Fox [32], Temple [118], and others, and have proved especially useful in engineering work. They are difficult to define precisely, since the computer is essentially advised to use all the art and artifice he can muster to

find x such that $r = b - Ax$ is near 0. Their predominant feature, however, is that the components of x are corrected, not in a predetermined order, but in an order of "worst first." If this feature is adopted as defining relaxation, the iteration function (defined in section 3) depends on x in a piecewise linear fashion, and the analytical character of the processes is completely different from that of the related linear processes. Relaxation has been studied recently in connection with solving systems of linear inequalities; see Agmon [2], and Motzkin and Schoenberg [83].

Other nonlinear iterative processes include the least-squares iterative methods discussed in sections 5, 6, and 7. They start with Cauchy [14], and are synthesized by Temple [118], Rosser [unpublished], Hestenes and Stein [53], and others. Special cases include certain linear processes which essentially deal with a positive definite matrix. For example, Kaczmarz [67] and Tompkins [120] interpret the system (1) as restricting $A^{-1}b$ simultaneously to n hyperplanes. A first guess $x_0$ is projected successively on each hyperplane in cyclic order. Then the distance $|x_k - A^{-1}b|$ decreases monotonically to 0. De la Garza [19] proposes a related process. When $A > 0$, the cyclic single-step method and all the relaxation methods are also least-squares methods. This fact is the basis of many studies of the relaxation methods; see [118] and Ostrowski [93].

Least-squares processes which are not linear include the gradient methods of [14] developed in [118], and by Kantorovič [68] and Birman [9].

see section 5. Their culmination appears to be the conjugate gradient method of Hestenes [50, 54], Lanczos [72], and Stiefel [114, 54]; this is a finite iteration described in section 7. A gradient method in a more general metric is mentioned in section 6.

Seemingly very different are the Monte Carlo methods for solving (1), employing random sampling of a certain chance variable whose expectation is $A^{-1}b$. One such method, devised by von Neumann and Ulam [unpublished], is described by Forsythe and Leibler [28]; Wasow [125] devises another Monte Carlo method. Both are based on properties of discrete random walks in a space of n points. When the system (1) represents Laplace's or related difference equations in one or more dimensions, the Monte Carlo methods have a longer history; see the exposition and bibliography in Curtiss [18]. The methods are theoretically fascinating, but there is little evidence yet of their practical utility for solving linear systems.

The iterative processes for solving (1) are likely to converge slowly, and a number of tricks have been devised to speed them up, called acceleration processes. Accelerations of linear processes may themselves be either linear or nonlinear; some are described in section 4. A number of processes for accelerating the nonlinear gradient methods are mentioned at the end of section 5, and the conjugate gradient method of section 7 may be considered also as an acceleration.

The distinction between direct and iterative methods is ordinarily stressed in numerical analysis; see [23], for example. Applied to systems (1), a direct method is one which yields $A^{-1}b$ exactly

after a finite number of arithmetical operations, if the latter are performed without round-off error. An iterative process can ordinarily yield $A^{-1}b$ only as the limit of a sequence of exact arithmetical operations. However, it must be remembered that as soon as calculations are rounded off (as ordinarily occurs in machine calculation), direct methods disappear except for quite simple problems, and all methods become iterative; see the end of section 8. The practical distinction between methods for solving (1) then appears to depend on: (a) the speed of the convergence to $A^{-1}b$, and (b) the simplicity of the computation at each stage of the iteration. A two-way classification of methods might well be based on their behavior with respect to properties (a) and (b), and that of [27] was roughly of this type. Two difficulties make such a classification imprecise. First, the theory of round-off error is too poorly developed to yield rates of convergence to $A^{-1}b$. Second, the practical criteria of simplicity in machine computation vary too greatly among different machines.

One may also distinguish whether the matrix A is altered in the course of solution, as in elimination, or whether it is retained in its original form, as in the conjugate gradient process. This is probably a crucial distinction in practice, for, when the original A is retained, the round-off errors seem to accumulate more slowly; see section 9.

3. Linear iterative processes. An iterative process for solving (1) (or other equation) is determined by the functions $F_k$ wherewith the $(k + 1)$-th approximant to $A^{-1}b$, $x_{k+1}$, is derived from the

earlier approximants $x_0, x_1, \cdots, x_k$. If the only argument of $F_k$ is $x_k$, the iterative process is said (Schröder [103]) to be of first degree: $x_{k+1} = F_k(x_k)$. If the function $F_k$ is independent of k, the process is called stationary. A bibliography on iteration as such has been prepared by Schwerdtfeger [106].

As elsewhere in mathematics, the most studied functions are the linear ones. We introduce the (most general) linear iterative process of the first degree by the definition

$$x_{k+1} = F_k(x_k) = H_k x_k + v_k \quad , \tag{2}$$

where the $H_k$ are square matrices and $v_k$ are vectors. If the iterative process described by (2) is to solve (1), it seems essential that the solution $A^{-1}b$ be a fixed point of $F_k$:

$$A^{-1}b = F_k(A^{-1}b) = H_k A^{-1}b + v_k \quad . \tag{3}$$

We demand that (3) hold. It follows that

$$v_k = (I - H_k) A^{-1}b = M_k b \quad . \tag{4}$$

If $H_k$ and $M_k$ are independent of b, then (6) follows from (4). Thus:

The most general stationary linear iterative process of the first degree for solving (1) which is independent of b and which satisfies (3) is defined by the functions

$$x_{k+1} = F_k(x_k) = H_k x_k + M_k b \quad , \tag{5}$$

where the square matrices $H_k$ and $M_k$ depend only on A and satisfy the relation

(6) $$H_k + M_k A = I \quad .$$

If the process is stationary, then

(7) $$x_{k+1} = Hx_k + Mb \quad ,$$

where the square matrices H and M depend only on A and satisfy the relation

(8) $$H + MA = I \quad .$$

Stationary processes of the type (7), (8) have been studied by Cesari [15], Wittmeyer [127], Geiringer [44], and others. They include the cyclic single-step iteration of Nekrasov [88] and Liebmann [74], and those of Jacobi [63], Richardson [98], Frankel [38], Young [129], and many others. In the cyclic single-step process, for example, one writes A as the sum of two matrices $B = (b_{ij})$ and $C = (c_{ij})$, where

$$b_{ij} = \begin{cases} a_{ij} & (i \geq j) \\ 0 & (i < j) \end{cases} \quad \text{and} \quad c_{ij} = \begin{cases} 0 & (i \geq j) \\ a_{ij} & (i < j) \end{cases} \quad .$$

Then $Bx_{k+1} + Cx_k = b$, or $x_{k+1} = -B^{-1}Cx_k + B^{-1}b$; it is assumed that no $a_{ii} = 0$. Thus $H = -B^{-1}C$, while $M = B^{-1}$.

Let us consider the convergence of the linear processes, under the assumption that all arithmetic operations are carried out with perfect accuracy. From (2) and (3) it follows that

(9) $$x_{k+1} - A^{-1}b = H_k(x_k - A^{-1}b) \quad ,$$

whence

$$x_k - A^{-1}b = K_k(x_o - A^{-1}b) \quad , \tag{10}$$

where $K_k = H_{k-1}H_{k-2} \cdots H_1H_o$. In order that $x_k - A^{-1}b \to 0$ for arbitrary $x_o$ it is therefore necessary and sufficient that

$$\lim_{k\to\infty} K_k z = 0 \quad , \quad \text{for every vector } z. \tag{11}$$

In practice condition (11) is usually known to hold only in certain special cases, such as when:

(a) all $H_k$ are polynomials in some one matrix - for example, A;

(b) all $H_k \equiv H$ (stationary process);

(c) with some norm function $\|\cdot\|$, all $\|H_k\| \leq 1 - \varepsilon < 1$.

Henceforward we consider only case (b): stationary linear processes. Then $K_k = H^k$, and it is known (see [94], for example) that (11) holds if and only if each eigenvalue $\lambda_i(H)$ of H is less than one in absolute value. Thus, we have derived the well known result that in the stationary linear process $x_k \to A^{-1}b$ for all $x_o$ if and only if all $|\lambda_i(H)| < 1$.

This result, while exceedingly important, hardly begins to solve the practical computer's problem. He raises the following questions: (i) How can one tell whether all $|\lambda_i(H)| < 1$ ? (ii) If all $|\lambda_i(H)| < 1$, how fast will the convergence be? (iii) If the convergence is intolerably slow (as the computer realistically expects), how may it be speeded up? To a considerable extent these questions

are answered by a knowledge of the asymptotic behavior of the error vector $x_k - A^{-1}b$ as $k \to \infty$. Question (iii) will also be dealt with in section 4.

The general theory of the Jordan canonical form [78, 94] of the matrix H can be applied to discuss the asymptotic behavior of $x_k - A^{-1}b$. Suppose, as is usually the case, that H has the diagonal canonical form

$$H = \sum_{i=1}^{n} \lambda_i c_i r_i^T \quad , \tag{12}$$

where $c_i$, $r_i^T$ are respectively the column and row eigenvectors of H belonging to $\lambda_i$. Suppose that $x_o - A^{-1}b = \sum_1^n \gamma_i c_i$. From (10) one then finds that

$$x_k - A^{-1}b = H^k(x_o - A^{-1}b) = \sum_{i=1}^{n} \gamma_i \lambda_i^k c_i \quad . \tag{13}$$

If $\lambda_n$ is a unique eigenvalue of maximum modulus, and if $\gamma_n \neq 0$, then

$$x_k - A^{-1}b \simeq \gamma_n \lambda_n^k c_n \quad (\text{as } k \to \infty) \quad . \tag{14}$$

This indicates that asymptotically $x_k \to A^{-1}b$ along a certain line, a fact which is often useful in accelerating the convergence of $\{x_k\}$.

When $\lambda_n$ is unique but $\gamma_n = 0$, the relation (14) does not hold for exact operations. However, a calculation involving round-off errors will soon have the practical effect of making $\gamma_n \neq 0$. If several eigenvalues dominate in absolute value, formula (14) must usually be modified to include several terms. In any case, when (12)

holds, formula (13) shows that $|x_k - A^{-1}b|$ approaches zero linearly — i.e., with the order of the k-th power of $\max_i |\lambda_i|$. When (12) fails (i.e., when H does not have a diagonal Jordan canonical form), formula (13) has to be altered. The important fact, however, is that the basic mathematical theory is already in existence for dealing with linear iterative processes — at least as long as round-off errors are not considered in detail. (It is commonly supposed that a non-diagonal canonical form would never arise in practical analysis. However, if one uses the Liebmann [74] process to solve the usual difference equation corresponding to the Dirichlet problem in one or more dimensions, the matrix H turns out to have a non-diagonal canonical form.)

Recent developments in stationary linear processes have included a suggestion of Aitken [4] and others for making $H > 0$ (i.e., a positive definite symmetric matrix), so that all $\lambda_i(H)$ will be real and positive, insuring that (13) will in fact be dominated by one term. In the cyclic single-step process this can be achieved by solving for the components in the "to-and-fro" order $1, 2, \cdots, n, n-1, \cdots, 2$ and repeat.

According to Ostrowski [91], for the cyclic single-step process it was first proved by Pizzetti [95] that, if $A > 0$, all $|\lambda_i(H)| < 1$. A converse recently proved by Reich [97], and more simply by Ostrowski [92], asserts that *if A is symmetric and each $a_{ii} > 0$, then all $|\lambda_i(H)| < 1$ if and only if $A > 0$*. Collatz [17], Stein and Rosenberg [113], and others have studied the relations between the $\lambda_i(H)$

for the cyclic single-step process and those for the related total-step process of Jacobi [63]. Other recent developments [38, 129] include the alteration of H by simple devices so that $\max_i |\lambda_i|$ is reduced as much as possible. In the Liebmann process for the Dirichlet problem, a certain systematic "over-relaxing" has just this effect.

4. Acceleration of stationary linear processes. Since ordinarily the stationary linear process (7) seems to converge slowly, or even diverge, it is of the greatest practical importance to devise methods to improve the convergence. Any procedure which replaces $x_k$ by a vector closer in some sense to $A^{-1}b$ is loosely called an acceleration of the iterative process (7). An acceleration is ordinarily a summability process applied to the sequence $\{x_k\}$ defined by (13). The best understood accelerations are, like the iterations themselves, linear processes, in which the improved vector is a linear combination of $x_0, x_1, \cdots, x_k$.

The simplest linear acceleration is useful when H has a unique dominant eigenvalue $\lambda_n$: By (14), for large k

$$x_{k+1} - A^{-1}b \doteq \lambda_n(x_k - A^{-1}b) \quad . \tag{15}$$

Hence the formula

$$A^{-1}b \doteq \frac{x_{k+1} - \lambda_n x_k}{1 - \lambda_n} \tag{16}$$

may be expected to give a better approximation to $A^{-1}b$ than $x_k$ gives, even when $\lambda_n$ is known only approximately. This idea has been applied by Flanders and Shortley [24], Lyusternik [77], and others.

The same idea can be extended to take care of two or more dominant $\lambda_i$. Let E be the operator increasing by one the subscript k of x: $x_{k+1} = Ex_k$. One can write (16) in the form

$$A^{-1}b \doteq P_1(E)\, x_k \quad , \tag{17}$$

where $P_1(E) = (1 - \lambda_n)^{-1} (E - \lambda_n)$ is a polynomial of degree one in E. Let m be an integer. Consider the formula

$$A^{-1}b \doteq P_m(E)\, x_k \quad , \tag{18}$$

where $P_m$ is some polynomial of degree m. How shall we determine $P_m$ so that (18) is a good approximation? If we can answer this effectively, then (18) is a useful linear acceleration of the process (7).

In order that the operation (18) not worsen an already good approximation $x_k$, we must have

$$P_m(1) = 1 \quad . \tag{19}$$

For if $x_k = A^{-1}b$, then, by (3), $x_{k+1} = x_{k+2} = \cdots = A^{-1}b$, so that $P_m(E)\, x_k = P_m(1)\, x_k = P_m(1)\, A^{-1}b$.

If $m = n$ = the order of H, and if $\phi(\lambda)$ is the <u>characteristic polynomial</u> of H, then the choice $P_m(E) = \phi(E)/\phi(1)$ is perfect, in that $P_m(E)\, x_k$ is exactly equal to $A^{-1}b$, for all $x_k$. But ordinarily $\phi$ is unknown, and to obtain it would involve precise knowledge of all the eigenvalues $\lambda_i$ of H.

Suppose that the eigenvalues $\lambda_i$, although not known precisely, are known all to be in a certain closed set R of the complex plane

(for example, a real interval). By (8), since $d(A) \neq 0$, the number 1 cannot be an eigenvalue of H. Hence we may assume that R is bounded away from the point 1. Now the eigenvalues of $P_m(H)$ are $\{P_m(\lambda_i)\}$. Since, by (13),

$$P_m(E)\, x_k - A^{-1}b = P_m(H)(x_k - A^{-1}b) = \sum_{i=1}^{n} P_m(\lambda_i)\gamma_i c_i \ , \qquad (10)$$

it is essential for the success of the acceleration that all $|P_m(\lambda_i)|$ be small. This leads to the approximation problem of determining the polynomial $P_m$ of degree m such that (19) holds and such that

$$\max_{\lambda \in R} |P_m(\lambda)| \text{ is a minimum} \ . \qquad (20)$$

Such polynomials, named generically after Čebyšev (= Chebyshev = Tschebyscheff = ···) arise frequently in numerical analysis.

If R is a real interval not containing 1, the minimizing polynomial $P_m(\lambda)$ is proportional to the ordinary Čebyšev polynomial $T_m(\lambda)$ for the interval R; see W. Markoff [79]. For just this reason the polynomials $T_m$ have been used several times in matrix problems to suppress the effects of unwanted eigenvalues; see Abramov [1], Flanders and Shortley [24], Gavurin [43], and Lanczos [73]. A recent review of Čebyšev approximation problems in the complex plane is in [80].

The real difficulty in numerical analysis is that R is not known. How can the information gained from computing $x_0, x_1, \cdots, x_k, \cdots, x_{k+m}$ be used effectively to localize R, so that the Čebyšev problem can be formulated? After R is localized, how can the corresponding

$P_m(E)$ be determined approximately? These are important questions which are in need of treatment. The use of symmetric matrices H helps a great deal, by restricting R to the real axis.

Another acceleration device is due to Schulz [104]. Suppose $A = I - H$, and that (21) converges. A linear process analogous to (7) obtains $A^{-1}$ as the sum of a Neumann series

$$(I - H)^{-1} = I + H + H^2 + \cdots \quad , \tag{21}$$

often slow to converge. By [104] one can obtain the partial sum $X_n = I + H + \cdots + H^{2^n-1}$ of (21) by n iterations (22),

$$X_{k+1} = X_k(2I - AX_k) \quad , \qquad X_o = I \quad , \tag{22}$$

a total of 2n matrix multiplications. The use of (22) to get $A^{-1}$ is sometimes called Newton's process. Can this idea be adapted to solving the system (1) without getting $A^{-1}$ first?

For accelerating the convergence of any linear process one also has the $\delta^2$-process of Aitken [3, 4] and its extensions by Shanks [108] and Samuelson [101]. Lubkin [76] has studied it as a nonlinear sequence-to-sequence summability process. It requires no knowledge of the $\lambda_i(H)$. Let $y_k$ represent an arbitrary, but fixed, component of $x_k$. Then the functional character of $y_k$ in a linear process is given by

$$y_k = y_\infty + \sum_{i=1}^{n} s_i \lambda_i^k \tag{23}$$

where $y_\infty$ is the desired component of $A^{-1}b$, and $\lambda_i$, $s_i$ are numbers.

To determine $y_\infty$ from (23) it is theoretically sufficient to have exactly $2n + 1$ successive values of $y_k$ — for instance, $y_0, y_1, \cdots, y_{2n}$. In practice the elimination of the $s_i$ and $\lambda_i$ would be too tedious, but frequently one $\lambda_i$, say $\lambda_n$, predominates in (23), and for moderately large k one may ignore $s_i$, $\lambda_i$ for $i = 1, \cdots, n-1$. In the $\delta^2$-process one assumes that, with sufficient accuracy, $y_k = y_\infty + s_n \lambda_n^k$, and calculates $y_\infty$ from three successive $y_k$. The formula is this:

$$y_\infty = \frac{y_k y_{k+2} - y_{k+1}^2}{y_k - 2y_{k+1} + y_{k+2}} = y_k - \frac{(\Delta y_k)^2}{\Delta^2 y_k} \quad .$$

In [108], m exponential terms are eliminated from (23) at once. If $m = 2$, for example, one can show that

$$y_\infty = \begin{vmatrix} y_k & \Delta y_k & \Delta^2 y_k \\ \Delta y_k & \Delta^2 y_k & \Delta^3 y_k \\ \Delta^2 y_k & \Delta^3 y_k & \Delta^4 y_k \end{vmatrix} \Bigg/ \begin{vmatrix} \Delta^2 y_k & \Delta^3 y_k \\ \Delta^3 y_k & \Delta^4 y_k \end{vmatrix} \quad .$$

5. Least-squares methods. A variational approach to solving (1) is very popular now. In a general treatment over the complex field (the present treatment is confined to the real field), Hestenes and Stein [53] take a matrix $R > 0$ and let $(z^T R z)^{\frac{1}{2}} = |z|_R$ be the R-length of a vector z. (The T stands for transposition.) If $B = A^T R A$ and $c = A^T R b$, then $B > 0$ also. Let the deviation of x from $A^{-1}b$ be measured by

$$f(x) = |Ax - b|_R^2 = |x - A^{-1}b|_B^2$$

(24)

$$= x^TBx - 2x^Tc + |b|_R^2 \quad .$$

Starting from an initial vector $x_o$, one can move x in various directions with the object of minimizing f(x). Clearly the minimum is attained just when $x = A^{-1}b$.

To simplify the exposition we now assume $A > 0$ and take $R = A^{-1}$. Then

(25) $$f(x) = x^TAx - 2x^Tb + b^TA^{-1}b \quad .$$

Although f(x) is not computable unless $A^{-1}b$ is known, it is sufficient in practice to minimize the computable function $f(x) - b^TA^{-1}b$. Fix x, and let $r = b - Ax$ be the residual of the system (1) at x. Let $d \neq 0$ determine a direction. Since $f(x + \alpha d) = \alpha^2 d^TAd - 2\alpha d^Tr + f(x)$, the value of the real parameter $\alpha$ for which $f(x + \alpha d)$ is a minimum is

(26) $$\alpha^* = d^Tr/d^TAd \quad ;$$

this $\alpha^*$ is called the optimum $\alpha$ (corresponding to x and d).

For any $\alpha = \beta\alpha^*$ one can compute $f(x + \alpha d)$ from the relation $f(x) - f(x + \alpha d) = (2\alpha\alpha^* - \alpha^2)d^TAd$, i.e., $f(x) - f(x + \beta\alpha^* d) = \beta(2 - \beta)\alpha^{*2}d^TAd$. Thus (when $d^Tr \neq 0$), $f(x + \beta\alpha^* d) < f(x)$ for $0 < \beta < 2$. The greatest reduction in f(x) comes when $\beta = 1$.

We now describe a general least-squares iterative process for solving (1). There must be prescribed: (i) a start $x_o$; (ii) a sequence of directions $\{d_k\}$; (iii) a sequence of ratios $\{\beta_k\}$. (In [53], but not here, R must also be prescribed.) For each

$k = 0, 1, 2, \cdots$, one determines $\alpha = \alpha_k^*$ by (26) so that $f(x_k + \alpha d_k)$ is minimized. Then one lets

$$x_{k+1} = x_k + \beta_k \alpha_k^* d_k \quad , \tag{27}$$

where $r_k = b - Ax_k$ and $\alpha_k^* = d_k^T r_k / d_k^T A d_k$. (It is not specified that the $\beta_k$ and the $d_k$ be determined a priori; they may depend on $x_k$.) If the sequences $\{\beta_k\}$ and $\{d_k\}$ satisfy (28) and (29) it is shown in [53] that, independently of $x_o$, $f(x_k) \to f(A^{-1}b)$, as $k \to \infty$, so that $x_k \to A^{-1}b$. The conditions are:

$$0 < \delta_1 \leq \beta_k \leq 2 - \delta_1 < 2 \qquad \text{(all k)} \quad ; \tag{28}$$

$$0 < \delta_2 = |d_k^T r_k| / (|d_k| \cdot |r_k|) \qquad \text{(all k)} \quad . \tag{29}$$

In [53] an alternate to (28) states that the $d_k$ recurrently span the space in a certain uniform manner.

Among others, the following two least-squares processes are also linear iterative processes: (a) the $d_k$ and $\beta_k$ are independent of the $x_k$; (b) $d_k = r_k$ but $\alpha_k = \beta_k \alpha_k^* = \alpha$ is a constant. When in (a) all $\beta_k = 1$ and the $d_k$ are the coordinate unit vectors in cyclic order, one has the cyclic single-step process of Nekrasov [88]. The use of $f(x)$ is very useful in studying any single-step process. In (b) one can write the linear iteration function of section 3 in the form $F(x) = x + \alpha(b - Ax) = (I - \alpha A)x + b$; the process is due to Richardson [98] and to von Mises and Pollazek-Geiringer [123], and converges whenever $0 < \alpha < 2/\max_i \lambda_i(A)$.

In the general case, however, the least-squares process is non-linear. When $d_k = r_k = -\frac{1}{2}\,\mathrm{grad}\, f(x_k)$, it is called a gradient method (or method of steepest descent). When $\beta_k = 1$, one has the optimum gradient method (since $f(x_k + \alpha d_k)$ is minimized as a function of $\alpha$), proposed by Cauchy [14] and studied by Temple [118], Kantorovič [68], Birman [9], and by Hestenes and Karush [52] for the eigenvalue problem. Some variations of the method are treated by Krasnosel'skiĭ and Krein [70a].

By (25) the surfaces $f(x) = \text{constant}$ are similar and similarly situated ellipsoids whose common center $A^{-1}b$ we seek. Any approximant $x_k$ lies on a certain ellipsoid $S_k$ of the family. The gradient, $-2r_k$, lies in the normal $\pi_1(x_k)$ to $S_k$ at $x_k$. Now $x_{k+1}$ is the unique point of $\pi_1(x_k)$ for which $f(x)$ is minimized. Since $f(x)$ is a quadratic function of distance along the normal, $x_{k+1}$ is located halfway between the two intersections of $\pi_1(x_k)$ with $S_k$. Moreover, $x_{k+1}$ is the point where $\pi_1(x_k)$ is tangent to an ellipsoid of the family.

Let $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of A. Kantorovič [68] shows that

$$(30) \qquad \left[\frac{f(x_{k+1})}{f(x_k)}\right]^{\frac{1}{2}} \leq \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} = \mu_1 < 1 \qquad (\text{all } x_k) \quad .$$

From this it follows that $f(x_k) \downarrow 0$, so that $x_k \to A^{-1}b$ and the process converges. As we mentioned in section 3, the practical computer's problem merely begins with such knowledge. Some experience convinces him that (a) $\mu_1$ is frequently very close to 1, and (b) after a few

steps $[f(x_{k+1})/f(x_k)]^{\frac{1}{2}}$ becomes and remains very close to $\mu_1$; see [25]. Observation (a) is borne out by the remark [6, p. 591] that, for certain matrices A of type $C^TC$, $\lambda_n/\lambda_1$ is likely to be of the order of $n^2$, whence $1 - \mu_1 \sim n^{-2}$. For finite-difference approximations to the Laplace operator over rectangular regions Frankel [38] shows that $1 - \mu_1 \sim n^{-1}$, where n is the number of lattice points. As to (b), for any A the value $\mu_1^2$ can always be attained by $f(x_{k+1})/f(x_k)$ when $x_k - A^{-1}b$ assumes certain directions in the plane spanned by $u_1$ and $u_n$ (defined below). From (a) and (b) we conclude that the optimum gradient method is frequently too slow for practical use.

As a guide to the possible acceleration of the method it would be valuable to know the asymptotic behavior of $x_k - A^{-1}b$, if arithmetical operations are done exactly. But, because $x_{k+1}$ is obtained from $x_k$ by a rational cubic transformation, theorems are hard to prove! It is a conjecture of Forsythe and Motzkin [30], proved only for $n = 3$ in [29], that in the optimum gradient method the error vector $x_k - A^{-1}b$ is asymptotically a linear combination of the eigenvectors $u_n$, $u_1$ of A belonging to the largest ($\lambda_n$) and least ($\lambda_1$) eigenvalue of A. (If there are eigenvectors of A orthogonal to $x_o - A^{-1}b$, one disregards the corresponding eigenvalues in determining $\lambda_1$ and $\lambda_n$.) A proof of the conjecture for $n \geq 4$ would be very desirable because, when the conjectured asymptotic relationship holds, for all sufficiently large k the points $x_k$, $x_{k+1}$, $x_{k+2}$, and $A^{-1}b$ are asymptotically coplanar. Thus one could accelerate the convergence of the optimum gradient method by occasionally minimizing

$f(x)$ in the plane through the endpoints of $x_k$, $x_{k+1}$, and $x_{k+2}$. The method has been used successfully in experiments [25] for $n = 6$ on an IBM Card-Programmed Calculator, where the average value of $f(x_{k+1})/f(x_k)$ over a series of about 100 steps was reduced from .9733 (optimum gradient method) to .6245 (optimum gradient method with acceleration every ninth step).

A second idea to speed up the optimum gradient method, the $\beta$-optimum method proposed by Hartree [46] and by Hestenes and Stein [53], is to give $\beta_k$ some constant value $\beta$ in the range $0 < \beta < 1$. In the test problem of [25], Stein [112] finds $\beta = 0.9$ to be approximately best, and with it the average value of $f(x_{k+1})/f(x_k)$ is found [25] to be .8204; in [112] .8065 is found for a shorter run. Although the convergence of the $\beta$-optimum method is slower (in this test) than that of the accelerated optimum gradient method, the former has the considerable advantage of being shorter to code for automatic machinery. The success of the $\beta$-optimum method is perhaps due to an inherent instability of the transformations. In 2 dimensions the transformation of $x_k = A^{-1}b$ to $x_{k+1} = A^{-1}b$ has no stable fixed directions when $\beta$ is slightly less than 1 [Motzkin and the author, unpublished].

6. General descent methods. The function $f$ of (24) has the following significant properties: $f(A^{-1}b) = 0$; $f(x) > 0$ if $x \neq A^{-1}b$; $f$ is a convex function of $x$. Any $f$ with these properties might serve as the basis for what Professor Motzkin calls a descent method. In such a method one has a first guess $x_0$, and sequences $\{d_k\}$ and $\{\beta_k\}$ as in section 5. As before, one finds $\alpha = \alpha^*_k$ minimizing $f(x_k + \alpha d_k)$, and selects $x_{k+1} = x_k + \beta_k \alpha^*_k d_k$.

Other suitable f would be

$$f(x) = \sum_{i=1}^{n} |r_i| \quad , \tag{31}$$

$$f(x) = \max_i |r_i| \quad . \tag{32}$$

Methods employing the latter norm functions f, with $d_k = -\operatorname{grad} f(x_k)$, are somewhat related to - but apparently do not include - the piecewise linear iterative processes. Agmon [2] discusses a relaxation method for linear inequalities from somewhat this point of view. Zuhovickiĭ [130] gives a gradient method in the metric (32) for minimizing f(x) for incompatible systems $Ax = b$.

7. Method of conjugate gradients. It is easy to show that the acceleration step of [25], discussed in section 5, is equivalent to finding the unique point $x'$ for which $f(x') = f(x_k + \alpha_o r_k + \alpha_1 A r_k)$ is minimized as a function of the two parameters $\alpha_o$ and $\alpha_1$. This suggests a generalization to p parameters discussed by Kantorovič [68] and Birman [9]. Let $x_o$ be any point, and let $r_o = b - Ax_o$. Define f(x) by (25). Let $\pi_p(x_o)$ be the p-space of points $x_o + \alpha_o r_o + \alpha_1 A r_o + \cdots + \alpha_{p-1} A^{p-1} r_o$, where $\alpha_o, \cdots, \alpha_{p-1}$ are real parameters. Let $x_1$ be the unique point in $\pi_p$ for which f(x) is minimized:

$$f(x_o + \alpha_o r_o + \cdots + \alpha_{p-1} A^{p-1} r_o) = \min. \tag{33}$$

To determine $x_1$ one has to solve a linear system of p equations in the p unknowns $\alpha_o, \cdots, \alpha_{p-1}$.

Define $\lambda_1, \cdots, \lambda_n$ as in section 5. Let the interval $\lambda_1 \leq \lambda \leq \lambda_n$ be transformed into the interval $-1 \leq t \leq 1$ by the affine transformation $t = \tau(\lambda)$, carrying $\lambda_n$ into -1, and $\lambda_1$ into 1. Then $\tau(0) = \delta = (\lambda_n + \lambda_1)(\lambda_n - \lambda_1)^{-1}$. Let $T_n(t)$ be the ordinary Čebyšev polynomial of degree n, normalized so that $T_n(1) = \max_{-1 \leq t \leq 1} |T_n(t)| = 1$. In these notations Birman [9] has proved that, for each A,

$$(34) \qquad \left[\frac{f(x_1)}{f(x_0)}\right]^{\frac{1}{2}} \leq \frac{1}{T_p(\delta)} = \mu_p < 1 \quad (\text{all } x_0) \ .$$

Thus (34) is the extension to $p > 1$ of (30).

Let $t_i = \tau(\lambda_i)$, so that $t_1 = 1$, $t_n = -1$. For a given A, the value $\mu_p$ cannot necessarily be attained by $[f(x_1)/f(x_0)]^{\frac{1}{2}}$. It will be attained, however, for any A with eigenvalues for which the corresponding $t_i$ include all the $p + 1$ values t with $|T_p(t)| = 1$. For such A, the $x_0$ for which $[f(x_1)/f(x_0)]^{\frac{1}{2}} = \mu_p$ are in the subspace spanned by just those eigenvectors $u_i$ belonging to the $\lambda_i$ for which $|T_p[\tau(\lambda_i)]| = 1$. The case $p = 1$ is exceptional, in that the maximum $\mu_1$ is always attained in (34) = (30), just because the two values $t = \pm 1$ where $|T_1(t)| = 1$ are necessarily $\tau$-images of eigenvalues.

For any fixed integer $p \geq 1$ the above p-step process can be iterated to yield a convergent procedure for finding $A^{-1}b$. Namely, for $k = 0, 1, \cdots$, one obtains $x_{k+1}$ as the unique point x in $\pi_p(x_k)$ for which $f(x)$ is minimized. The optimum gradient method of section 5 is the case $p = 1$. If $p << n$ we may expect that there are eigenvalues $\lambda_i$ of A close to the $p + 1$ points where

$|T_p[\tau(\lambda_i)]| = 1$, and hence that the value $\mu_p$ of (34) is almost attained for certain $x_o$. It is then to be expected that, for most $x_o$, $[f(x_{k+1})/f(x_k)]^{\frac{1}{2}}$ will be approximately $\mu_p$ for all large k. Moreover, if $\lambda_1 << \lambda_n$, then $\delta$ is near 1 and $\mu_p \doteq 1 - 2p^2(\lambda_1/\lambda_n)$. Thus the minimization in p dimensions may be expected asymptotically to proceed $p^2$ times as fast as the optimum gradient method (p = 1), when $p << n$. When $\lambda_n/\lambda_1 \doteq n^2$, as considered in section 5, the iterated p-dimensional minimization in n-space may be expected to converge like the optimum gradient method in n/p dimensions.

The true asymptotic behavior of $x_k - A^{-1}b$ is unknown. Does the vector $x_k - A^{-1}b$ asymptotically lie in a certain (p + 1)-dimensional subspace, as is conjectured for p = 1?

Because the above iterative process requires at each step solving a linear system of p-th order, up to very recently the method was considered practical only for $p << n$. For p = n, in particular, it appeared that determining the minimizing $\alpha_i$ in (33) would involve solving a linear system quite as difficult as Ax = b. Then, about 1951, Stiefel [114], Lanczos [72], and Hestenes [50], working independently at first, all discovered that (33) can be minimized by p repetitions of a beautifully simple algorithm. By taking p = n one actually finds $A^{-1}b$ in n steps, except for round-off errors. The resulting conjugate gradient method is thus a typical finite iteration. An extended exposition is in [54], while brief expositions are given by Rosser [100], Householder [59], and Taussky and Todd [117].

The conjugate gradient method is a nonlinear stationary iterative process. The first approximant, $x_o$, is arbitrary; one takes $p_o = r_o = b - Ax_o$. For any $k \geq 0$, assume the vectors $x_k$, $r_k$, $p_k$ have been determined. Then $x_{k+1}$, $r_{k+1}$, $p_{k+1}$ are determined in order as follows:

$$
(35) \qquad \begin{cases} \alpha_k = r_k^T p_k / p_k^T A p_k \ ; \\ x_{k+1} = x_k + \alpha_k p_k \ ; \\ r_{k+1} = r_k - \alpha_k A p_k \ ; \\ \beta_k = -r_{k+1}^T A p_k / p_k^T A p_k \ ; \\ p_{k+1} = r_{k+1} + \beta_k p_k \ . \end{cases}
$$

Here $r_k = b - Ax_k$, and the significance of $p_k$, $\alpha_k$, $\beta_k$ will appear below. In the absence of round-off errors $x_n = A^{-1}b$; if round-off errors make $x_n \neq A^{-1}b$, one has merely to carry the algorithm on for $k = n + 1, n + 2, \cdots$, until sufficient accuracy is attained.

The kernel of a number of methods of solving $Ax = b$ for $A > 0$ is the determination of a set of n directions $\{p_k\}$ $(k = 0, \cdots, n - 1)$ which are conjugate (A-orthogonal) in the sense that $p_i^T A p_j = 0$ for $i \neq j$. If the $\{p_i\}$ are known, then

$$
(36) \qquad A^{-1}b = \sum_{k=0}^{n-1} (p_k^T b / p_k^T A p_k) p_k \ .
$$

A convenient method to get the $p_k$ is to apply the Gram-Schmidt process [99] of successively orthogonalizing some set of n linearly

independent vectors $\{v_k\}$. In Gaussian elimination (pivotal condensation) the $v_k$ are taken to be the n coordinate unit vectors, as Fox, Huskey, and Wilkinson [35] discovered, and the coefficients defining the orthogonalization build up a triangular matrix.

In the method of conjugate gradients the $v_k$ are the vectors $r_k = b - Ax_k$ $(k = 0, \cdots, r-1)$. The beauty of this choice is that $r_{k+1}$ turns out to be automatically conjugate to $p_0, p_1, \cdots, p_{k-1}$. In picking $p_{k+1}$ it is therefore necessary only to alter $r_{k+1}$ in the direction $p_k$; the calculation of $\beta_k$ and $p_{k+1}$ in (35) has this object. The calculation of $\alpha_k$ and $x_k$ is an iterative procedure for building up the sum (36).

Recalling that $x_p$ actually minimizes (33), we see that in practice the residual $r_p$ may become so small for some $p < n$ that it is unnecessary even to complete the n steps theoretically required to compute $A^{-1}b$. This occurred for $p = 90$ in a successful calculation with $n = 105$ of a difficult stress problem reported by Hochstrasser [56] and Stiefel [114]. Such a saving could hardly occur with Gaussian elimination, since the unit vectors have no such intimate connection with the system (1) as do the vectors $r_k$.

The conjugate gradient method has a geometrical interpretation. As in section 5, one seeks the common center $A^{-1}b$ of the ellipsoids $f(x) = \text{constant}$. Let $x_o$ be prescribed. One gets $x_1$ by performing one optimum gradient step (section 5) in the unrestricted n-dimensional space $R_n$. Recall the definition of $\pi_1(x_o)$. There is an $(n-1)$-dimensional affine subspace $R_{n-1}$ passing through $x_1$ and conjugate to $\pi_1(x_o)$.

The solution $A^{-1}b$ lies in $R_{n-1}$. One gets $x_2$ by taking an optimum gradient step within $R_{n-1}$. The gradient $p_1$ of $f(x)$ at $x_1$ within $R_{n-1}$ is the projection into $R_{n-1}$ of the gradient $r_1$ of $f(x)$ at $x_1$ within $R_n$. The optimal point in the gradient direction $p_1$ from $x_1$ is $x_2 = x_1 + \alpha_1 p_1$. Similarly, one gets $x_{k+1}$ by taking one optimum gradient step from $x_k$ within the $(n - k)$-dimensional affine subspace $R_{n-k}$ through $x_k$ conjugate to $\pi_k(x_o)$. Finally, $R_o$ is the solution point $A^{-1}b$.

This is a bare description of the method. In [54] Hestenes and Stiefel give an amazing number of its properties, discuss its application to unsymmetrical systems (1), and so on. A few machine experiments [51, 56, 57, 114] with the method suggest good stability with respect to round-off errors, but a theoretical study of the stability would be desirable.

The conjugate gradient method can theoretically be applied to solving a system $Ax = b$, where A is a bounded positive-definite self-adjoint operator on a Hilbert space. One defines $f(x) = (x, Ax) - 2(x, b) + (b, A^{-1}b)$; the inverse operator $A^{-1}$ certainly exists. The method will ordinarily no longer converge in n steps, but the asymptotic behavior of $f(x_p)$ can be discussed. Karush [70] shows that if $A = \alpha I + K$, where K is completely continuous and $\alpha \neq 0$, then $f(x_p)$ goes to 0 faster than the p-th term of any geometrical progression. Hayes [48] treats a general A with lower, upper bounds $m, M$ $(0 < m < M < \infty)$, and proves that $f(x_p) \leq [1 - (m/M)]^p f(x_o)$. More can be proved:

Let $\delta = (M + m)(M - m)^{-1}$. The Birman inequality (34) shows that $[f(x_p)/f(x_o)]^{\frac{1}{2}} \leq 1/T_p(\delta) < 1$. Hence, since $2T_p(\delta) = [\delta + \sqrt{\delta^2 - 1}]^p + [\delta - \sqrt{\delta^2 - 1}]^p$, one gets the estimate

$$(37) \qquad \left[\frac{f(x_p)}{f(x_o)}\right]^{\frac{1}{2p}} \leq \frac{2^{1/p}}{\delta + \sqrt{\delta^2 - 1}} \to \frac{1}{\delta + \sqrt{\delta^2 - 1}} \quad .$$

In another paper the asymptotic nature of $[f(x_p)]^{1/2p}$ will be described more precisely for a class of operators A with a continuous spectrum.

8. Errors and "condition." One must say something about the important but little-understood subject of errors. We may distinguish between: I. errors committed in the course of solving the system by a specific algorithm; II. errors inherent in the system $Ax = b$.

Within I one is concerned with the truncation errors and the round-off errors of an algorithm, a distinction explained in [124]. The truncation error exists for infinite iterations, and may be identified with $x_k - A^{-1}b$; its behavior has been examined in the above survey, under the assumption that there was no round-off error. The study of round-off error itself is far more difficult, and there seems to have been a complete discussion in connection with only one method, elimination; see von Neumann and Goldstine [124, 45], and also Mulholland [85]. For other studies see Bargmann, Montgomery, and von Neumann [6], Dwyer [20], Satterthwaite [102], Tuckerman [121], and Turing [122].

Any approximate solution $\xi$ of $Ax = b$ can be checked *a posteriori* by forming the residual $\rho = b - A\xi$. The magnitude of $A^{-1}b - \xi$ can then be estimated by using some tool for examining errors under II. Hence to bound the errors in a calculated $\xi$ it is unnecessary to have an *a priori* knowledge of the accumulation of round-off error. Such knowledge may be important, however, for planning purposes - for example, in deciding in advance how many digital places to carry in order that $\rho$ be reasonably small.

The errors under II have attracted more study. The practical analyst, realizing that the elements of A and b are subject to uncertainty, wishes to know the corresponding uncertainty in $A^{-1}b$; the latter is the *inherent error* of Milne [81]. The usual approach is the approximate one of bounding the principal part of the error, $\delta(A^{-1}b) = A^{-1}(\delta A)A^{-1}b + A^{-1}\delta b$; see Blumenthal [11], Milne [81], Moulton [84], Ostrowski [90], Wittmeyer [126], and Zurmühl [131]. But others (e.g., Janet [64], Ostrowski [89], Lonseth [77]) bound the entire error, finding a region S to which $A^{-1}b$ is rigorously confined, given $\rho = b - A\xi$ and other reasonably computable quantities associated with the system. See also Woodbury [128].

Various persons (e.g., Jürgens [66], Todd [119], and Turing [122]) have attempted to ascribe a *condition* to the system $Ax = b$. In part, the condition should measure the influence on $A^{-1}b$ of small changes in A and b; the larger the change in $A^{-1}b$ for given changes in A and b, the "worse" the condition. Although the condition depends on both A and b, the measures hitherto proposed depend only on A. The

belief is widespread that the condition of a system (1) has a decisive influence on the convergence of an iterative solution and on the accuracy of a direct solution; this cannot always be true. Even when it is true for an iterative process, it may be possible actually to take advantage of the poor condition of (1) in converting the slow process into an accelerated method which converges rapidly. There is great need for clarification of the group of ideas associated with "condition."

With the concept of "ill-conditioned" systems $Ax = b$ goes the idea of "pre-conditioning" them. Gauss [41] and Jacobi [63] made early contributions to this subject. That of Gauss is analyzed and extended in [31].

A convenient means of pre-conditioning is to premultiply the system with a matrix B, so that one has to solve

$$BAx = Bb \quad . \tag{38}$$

The perfect choice of B would be $A^{-1}$. A frequent choice is $A^T$, so that (38) gets a symmetric matrix $A^TA$, very convenient for many processes, though "worse conditioned" in some senses (Taussky [115]). Is there in any sense a "best choice" of B which is quickly obtainable?

Gauss' elimination process may be written in the form (38), where in the absence of round-off errors BA is a triangular matrix. In some calculations the true solution $A^{-1}b$ comes from iteration of the "back solution" - i.e., of getting $x_{k+1}$ by an approximate solution of the triangular system $BA(x - x_k) = B(b - Ax_k)$. Where this occurs, we may interpret the "forward solution" or triangularization as merely a preconditioning of the system (1) into the form (38).

9. Influence of computing equipment. The usefulness of a process for solving (1) depends intimately on the properties of the machine on which the calculation takes place, as well as on the special character of A and b. The past decade has seen revolutionary developments in computing equipment: analogue machinery, desk computers, IBM equipment and automatically sequenced high-speed digital computers. As a result, computing methods are in no way settled down, and bibliographies are out of date before publication.

Analogue machinery can be very useful, but is not discussed here; for references see Engineering Research Associates [22], Frame [36], and Murray [86].

While direct methods for solving Ax = b have been little mentioned here, they have been very successful since the time of Gauss or earlier. Dwyer [20], Bodewig [12] and others conclude that a compact arrangement of Gaussian elimination is commonly the best method for the computing team of a desk machine, a data sheet, and a trained human being - principally because the number of operations is minimized. Elimination is very successful with IBM machines also, but its superiority over other methods is less pronounced, because it is seldom expedient with IBM equipment to use the compact arrangements which save so many operations. A bibliography on computing methods for IBM machinery is given in [38a].

Let us now consider automatically programmed digital computers like the SWAC. These are much faster than previous computers, but the speed of arithmetic operations and of access to a small store of data

(high-speed memory) has been accelerated a great deal more than the operations of input, output, and access to a large store of data. The resulting change in the relative costs of different operations has a profound effect on the choice of computing methods. One soon learns that a variety of processes have been tried and advocated for solving (1); certainly the optimal method depends on the problem, the machine, the operator, and the coder. Moreover, small changes in these factors may radically alter the optimal choice of method.

The following tentative ideas are based on a limited experience with the SWAC, and practically no experience with other machines. (The analysis is dominated by the relative shortage of memory cells in the SWAC; it is therefore less pertinent for machines with more storage space, and for the SWAC after the expected addition of a magnetic drum.) Assume n to be fairly large - say $\geq 15$. For simplicity we again confine our attention to matrices $A > 0$. As indicated after (38), the forward solution in elimination amounts to building up a new matrix BA, which must be stored somewhere. If BA is kept in the high-speed memory, it occupies critically needed space. If it is output and input as needed (say by rows, as Huskey [61] describes), the programming is complicated and the solution is considerably slowed. If A is a matrix which can be generated internally as needed (for instance, the matrix of the Laplace difference operator), it requires little space, and BA becomes the principal item to be stored. Where A cannot be generated internally, the storage problem gets still worse, because the round-off errors can only be reduced by

using A to compute the residual $r = b - Ax$ from time to time, so that both BA and A must be stored.

These considerations suggest that a solution process should preferably work with only one matrix, A itself, and should require relatively little other storage. Since the instructions have to be stored, this suggests a process of simple structure, repeated as often as necessary. A process seeming to require a minimum of storage is the cyclic single-step procedure mentioned in section 3; besides A (if it must be stored), one need store only one vector, x, at a time. This method was picked by Reich [96] as best for a digital computer, and is undoubtedly ideal when it converges fast enough. But we may expect that the convergence is frequently too slow. If an acceleration of the types discussed in section 4 is needed, the complication in programming may make another procedure preferable. Another method of speeding up the cyclic single-step method is by appropriately overcorrecting (or undercorrecting ?) at each step, as discussed by Frankel [38] and Young [129] for special systems (1). It seems likely that a careful determination of the optimal overcorrection will sometimes provide adequate convergence, but that it will often fail.

The ordinary relaxation (i.e., piecewise linear) processes require about the same storage as the cyclic single-step methods; it is not clear whether they are essentially faster or not. A suggestion of Motzkin and Schoenberg [83] for extreme overrelaxation is promising but untried.

If the above methods fail, one can switch to the optimum gradient method of section 5. This also works with A, which must be stored or generated, and further requires the storage of the two vectors $x_k$ and $r_k$. (The storage of $x_k$ can be avoided if $x_{k+1} - x_k$ is output at each step, and cumulated later.) Again the method is probably commonly too slow. It can be speeded up either by the $\beta$-optimum device of section 5, for $\beta < 1$, or by Richardson's idea [98] of widely varying $\alpha_k = \beta\alpha_k^*$ over the range of eigenvalues $\lambda_i^{-1}$ of $A^{-1}$.

If these tricks fail or require too complex a program, the gradient methods of section 7 are available. Besides A, they require the storage of three vectors $x_k$, $r_k$, $p_k$. (As above, the outputting of $x_{k+1} - x_k$ saves storing $x_k$.) Of these methods, there seems to be no reason for not adopting the optimum gradient method, since for the same storage its convergence is much the best. Programming is simple, as only one routine is needed; all necessary variations in the $\alpha_k$ are provided automatically. A drawback is that, since the $\alpha_k$ and other numbers vary so much in the calculation, it is difficult to provide scaling factors in advance. Consequently one uses "floating binary point" operations, requiring considerable memory space to hold the instructions and multiplying the solution time by a factor which varies on the SWAC from the order of 50 (when A is generated internally) to about one (when A is input at each iterative step). But the method has proved able to cope with some very "badly conditioned" matrices, as reported by Hestenes, Hochstrasser, and Wilson [51], and Hochstrasser [57]. It probably approaches the

ideal of a machine method which can be relied on to work automatically without special analysis of the particular system (1).

With any method the partitioning of A may greatly increase the speed by enabling subsidiary matrix inversions to take place entirely within the high-speed memory; see [21]. One usually thinks of Gaussian elimination on the submatrix blocks. Would other methods on the blocks be preferable?

BIBLIOGRAPHY

This is a selection of 133 titles out of over 500 in my card file. It is obvious that many valuable references have had to be omitted. Ordinarily, journal names are abbreviated according to the list in Mathematical Reviews vol. 11 (1950) pp. 859-872. Russian is transliterated according to the system loc. cit., p. 873.

1. A. A. Abramov, On a method of acceleration of iterative processes (Russian), Doklady Akad. Nauk SSSR vol. 74 (1950) pp. 1051-1052.

2. Shmuel Agmon, The relaxation method for linear inequalities, NAML Report 52-27, National Bureau of Standards, Los Angeles, 1951, multilithed typescript, 23 pp. To appear in Ann. of Math. (1)

3. A. C. Aitken, On Bernoulli's numerical solution of algebraic equations, Proc. Roy. Soc. Edinburgh. vol. 46 (1926) pp. 289-305.

4. _____, Studies in practical mathematics. V. On the iterative solution of a system of linear equations, Proc. Roy. Soc. Edinburgh. Sect. A. vol. 63 (1950) pp. 52-60.

D. G. Aronson: see [26].

5. T. Banachiewicz, Méthode de résolution numérique des équations linéaires, du calcul des déterminants et des inverses, et de réduction des formes quadratiques Bull. Int. Acad. Polon. Sci. Cl. Sci. Math. Nat. Ser. A. Sci. Math. (1938) pp. 393-404. Also in Cracow observatory reprint 22.

6. V. Bargmann, D. Montgomery, and J. von Neumann, Solution of linear systems of high order, Report prepared for the Bureau of Ordnance (Contract NORD-9596), 25 October 1946, multigraphed, bound, 86 pp.

7. Commandant Benoit, Note sur une méthode de résolution des équations normales provenant de l'application de la méthode des moindres carrés à un système d'équations linéaires en nombre inférieur à celui des inconnues.—Application de la méthode à la résolution d'un système défini d'équations linéaires (Procédé du Commandant Cholesky), Bull. Geodésique year 1924 (1924) no..2 pp. 67-77.

8. M. D. Bingham, A new method for obtaining the inverse matrix, J. Amer. Statist. Assoc. vol. 36 (1941) pp. 530-534.

9. M. Š. Birman, Some estimates for the method of steepest descent (Russian), Uspehi Matem. Nauk (N.S.) vol. 5 (1950) no. 3 pp. 152-155. Translated by C. D. Benster in NBS Report 2007, National Bureau of Standards, Los Angeles, August, 1952.

10. A. N. Black and R. V. Southwell, Relaxation methods applied to engineering problems. II. Basic theory, with application to surveying and to electrical networks and an extension to gyrostatic systems, Proc. Roy. Soc. London. Ser. A. vol. 164 (1938) pp. 447-467.

11. O. Blumenthal, Über die Genauigkeit der Wurzeln linearer Gleichungen, Zeitschrift für Math. und Physik vol. 62 (1914) pp. 359-362.

12. E. Bodewig, Bericht über die verschiedenen Methoden zur Lösung eines Systems linearer Gleichungen mit reellen Koeffizienten. I, II, III, IV, V, Nederl. Akad. Wetensch., Proc. vol. 50 (1947) pp. 930-941, 1104-1116, 1285-1295, and vol. 51 (1948). pp. 53-64, and 211-219 = Nederl. Akad. Wetensch. Indagationes Math. vol. 9 (1947). pp. 441-452, 518-530, 611-621, and vol. 10 (1948). pp. 24-35 and 82-90.

13. H. Boltz, Entwicklungsverfahren zur Ausgleichung geodätischer Netze nach der Methode der kleinsten Quadrate, Veröffentlichen des Preussischen Geodätischen Instituts N. F. no. 90 (1923), Berlin.

14. A. L. Cauchy, Méthode générale pour la résolution des systèmes d'équations simultanées, C. R. Acad. Sci. Paris vol. 25 (1847), pp. 536-538.

15. Lamberto Cesari, Sulla risoluzioni dei sistemi di equazioni lineari per approssimazioni successive, Extract from Rassegna delle poste, dei telegrafi e dei telefoni vol. 4 (1937), 37 pp.

A. R. Collar: see [39].

16. L. Collatz, Graphische und numerische Verfahren, pp. 1-92 of Alwin Walther, Applied mathematics. Part I, FIAT Review of German Science 1939-1946, Office of Military Government, Wiesbaden, 1948, 307 pp.

17. ————, Über die Konvergenzkriterien bei Iterationsverfahren für lineare Gleichungssysteme, Math. Z. vol. 53 (1950) pp. 149-161.

18. John H. Curtiss, Sampling methods applied to differential and difference equations, pp. 87-109 of Proceedings, seminar on scientific computation November, 1949, International Business Machines Corp., New York, 1950, 109 pp.

19. A. de la Garza, An iterative method for solving systems of linear equations, Report K-731, Carbide and Carbon Chem. Div., Union Carbide and Carbon Corp., K-25 Plant, Oak Ridge, Tenn., 15 Feb. 1951, mimeographed, 15 pp.

W. J. Duncan: see [39].

20. Paul S. Dwyer, Linear computations, Wiley, New York, 1951, 344 pp.

21. Eckert-Mauchly Div. of Remington-Rand Corp., Matrix algebra programs for the UNIVAC, Philadelphia, 1951, typescript, 11 pp.

22. Engineering Research Associates, High-speed computing devices, McGraw-Hill, New York, 1950, 451 pp. (Supervisors: C. B. Tompkins and J. H. Wakelin; editor: W. W. Stifler, Jr.)

23. V. N. Faddeeva, Вычислительные методы линейной алгебры (Computational methods of linear algebra; Russian), Moscow-Leningrad, 1950, 240 pp. (A translation is being prepared at the National Bureau of Standards, Los Angeles.)

24. Donald A. Flanders and George Shortley, Numerical determination of fundamental modes, J. Appl. Phys. vol. 21 (1950) pp. 1326-1332.

25. A. I. and G. E. Forsythe, Punched-card experiments with accelerated gradient methods for linear equations, NBS Report 1643, National Bureau of Standards, Los Angeles, March 1952, multilithed typescript, 29 pp. To appear in J. Research Nat. Bur. Standards.

26. George E. Forsythe, Theory of selected methods of finite matrix inversion and decomposition, Lecture notes by D. G. Aronson and K. E. Iverson, INA Report 52-5, National Bureau of Standards, Los Angeles, 1951, multilithed typescript, 93 pp.

27. _____, Tentative classification of methods and bibliography on solving systems of linear equations, INA Report 52-7, National Bureau of Standards, Los Angeles, 1951, multilithed typescript, 78 pp. To appear in [87].

28. George E. Forsythe and Richard A. Leibler, Matrix inversion by a Monte Carlo method, Math. Tables and Other Aids to Computation vol. 4 (1950) pp. 127-129 and vol. 5 (1951) p. 55.

29. G. E. Forsythe and T. S. Motzkin, Asymptotic properties of the optimum gradient method, abstract, Bull. Amer. Math. Soc. vol. 57 (1951) p. 183.

30. _____, Acceleration of the optimum gradient method. Preliminary report, abstract, Bull. Amer. Math. Soc. vol. 57 (1951) pp. 304-305.

31. _____, An extension of Gauss' transformation for improving the condition of systems of linear equations, Math. Tables and Other Aids to Computation vol. 6 (1952) pp. 9-17.

32. L. Fox, A short account of relaxation methods, Quart. J. Mech. Appl. Math. vol. 1 (1948) pp. 253-280.

33. ———, Practical solution of linear equations and inversion of matrices, manuscript, about 60 pp. To appear in J. Research Nat. Bur. Standards.

34. ———, Practical methods for the solution of linear equations and the inversion of matrices, J. Roy. Statist. Soc. Ser. B. vol. 12 (1950) pp. 120-136.

35. L. Fox, H. D. Huskey, and J. H. Wilkinson, Notes on the solution of algebraic linear simultaneous equations, Quart. J. Mech. Appl. Math. vol. 1 (1948) pp. 149-173.

36. J. S. Frame, Machines for solving algebraic equations, Math. Tables and Other Aids to Computation vol. 1 (1945) pp. 337-353.

37. [Stanley P. Frankel], Bibliography on computing machines, (4) Analysis Laboratory, Calif. Inst. of Technology, Pasadena, October 1949, hectographed typescript, 44 pp. (Frankel's name not on copy.)

38. Stanley P. Frankel, Convergence rates of iterative treatments of partial differential equations, Math. Tables and Other Aids to Computation vol. 4 (1950) pp. 65-75.

38a. Ann D. Franklin and Eric V. Hankam, Bibliography on the use of IBM machines in science, statistics, and education, International (4) Business Machines Corp., New York, 1952, 50 pp. See also [62].

39. R. A. Frazer, W. J. Duncan, and A. R. Collar, Elementary matrices and some applications to dynamics and differential equations, Cambridge Univ. Press, 1938, 416 pp.

40. Konrad Friedrich and Werner Jenne, Geometrisch-anschauliche Auflösung linearer mit Nullkoeffizienten ausgestatteter Gleichungssysteme, Deutsche Akad. Wiss. Berlin. Veröff. Geodät. Inst. Potsdam, no. 5 (1951), 68 pp.

41. C. F. Gauss, Letter to Gerling, 26 December 1823, Werke, Göttingen, vol. 9 pp. 278-281. Translated by G. E. Forsythe, Math. Tables and Other Aids to Computation vol. 5 (1951) pp. 255-258.

42. _____, Supplementum theoriae combinationis observationum erroribus minimis obnoxiae, 1826, Werke, Göttingen, vol. 4 pp. 55-93.

43. M. K. Gavurin, Application of polynomials of best approximation to improving the convergence of iterative processes (Russian), Uspehi Matem. Nauk (N.S.) vol. 5 (1950) no. 3 pp. 156-160. Translated by C. D. Benster in NBS Report 2007, National Bureau of Standards, Los Angeles, August 1952.

44. Hilda Geiringer, On the solution of systems of linear equations by certain iteration methods, pp. 365-393 of Reissner Anniversary Volume, Contributions to applied mechanics, Edwards, Ann Arbor, Mich., 1949.

_____: see also [123].

45. Herman H. Goldstine and John von Neumann, Numerical inverting of matrices of high order, II, Proc. Amer. Math. Soc. vol. 2 (1951) pp. 188-202. Part I is [124].

Eric V. Hankam: see [38a].

46. D. R. Hartree, Experimental arithmetic, Eureka vol. 10 (1948) pp. 13-18.

47. Harvard Computation Laboratory, A manual of operation for the automatic sequence controlled calculator, Harvard Univ. Press, 1946, 561 pp.

48. R. M. Hayes, Iterative methods of solving linear problems on Hilbert space, Ph.D. Thesis, Univ. of California, Los Angeles, 1952, multilithed typescript, 62 pp.

49. A. Hertwig, Die Lösung linearer Gleichungen durch unendliche Reihen und ihre Anwendungen auf die Berechnung hochgradig statisch unbestimmter Systeme, pp. 37-59 of Festschrift für H. Müller-Breslau, Leipzig, 1912.

50. M. R. Hestenes, Iterative methods for solving linear equations, NAML Report 52-9, National Bureau of Standards, Los Angeles, 1951, multilithed typescript, 19 pp.

51. M. R. Hestenes, Urs Hochstrasser, and L. S. Wilson, Some numerical examples on solving systems of linear equations by the conjugate gradient method for non-symmetric systems of equations, National Bureau of Standards, Los Angeles, manuscript.

52. Magnus R. Hestenes and William Karush, A method of gradients for the calculation of the characteristic roots and vectors of a real symmetric matrix, J. Research Nat. Bur. Standards vol. 47 (1951) pp. 45-61.

53. Magnus R. Hestenes and Marvin L. Stein, The solution of linear equations by minimization, NAML Report 52-45, National Bureau of Standards, Los Angeles, 12 December 1951, multilithed typescript, 35 pp.

54. Magnus R. Hestenes and Eduard Stiefel, Method of conjugate gradients for solving linear systems, NBS Report 1659, National Bureau of Standards, Los Angeles, 10 March 1952, multilithed typescript, 101 pp. To appear in J. Research Nat. Bur. of Standards.

55. Thomas J. Higgins, A survey of the approximate solution of two-dimensional physical problems by variational methods and finite difference procedures, pp. 169-198 of L. E. Grinter (editor), Numerical methods of analysis in engineering, MacMillan, New York, 1949, 207 pp.

56. [Urs] Hochstrasser, Die Berechnung der Verschiebungen in einer parallelogrammförmigen Scheibe, Bericht S-17, Techn. Abteilung, Eidgenössisches Flugzeugwerk, Emmen, Switzerland, multigraphed typescript, 21 pp.

57. Urs Hochstrasser, The solution of systems of linear equations by the conjugate gradient method for use on IBM equipment, National Bureau of Standards, Los Angeles, manuscript.

___ _____: see also [51].

58. Harold Hotelling, Some new methods in matrix calculation, Ann. Math. Statistics vol. 14 (1943) pp. 1-33 and 440-441.

59. A. S. Householder, Notes on numerical methods, Oak Ridge National Laboratory, Tenn., 1949-1952, multilithed typescript, c. 400 pages so far.

60. _____, Some numerical methods for solving systems of linear equations, Amer. Math. Monthly vol. 57 (1950), pp. 453-459.

61. H. D. Huskey, Characteristics of the Institute for Numerical Analysis computer, Math. Tables and Other Aids to Computation vol. 4 (1950) pp. 103-108.

_____: see also [35].

62. International Business Machines Corp., Bibliography on the use of IBM machines in science, statistics, and education, New York, 1950. Apparently superseded by [38a].

K. E. Iverson: see [26].

63. C. G. J. Jacobi, Ueber eine neue Auflösungsart der bei der Methode der kleinsten Quadrate vorkommenden linearen Gleichungen, Astr. Nachr. vol. 22 (1845) no. 523 pp. 297-306. (Also in Jacobi's Werke vol. 3, p. 467?)

64. M. Janet, Sur les systèmes d'équations aux dérivées partielles, C. R. Acad. Sci. Paris vol. 170 (1920) pp. 1101-1103 and J. Math. Pures Appl. Sér. 8, vol. 3 (1920), pp. 65-151.

Werner Jenne: see [40].

65. Henry Jensen, An attempt at a systematic classification of some methods for the solution of normal equations, Meddelse No. 18, Geodaetisk Institut, Copenhagen, 1944, 45 pp.

66. Enno Jürgens, Zur Auflösung linearer Gleichungssysteme und numerischen Berechnung von Determinanten, Palm (printer), Aachen, 1886, 20 pp.

67. S. Kaczmarz, Angenäherte Auflösung von Systemen linearer Gleichungen, Bull. Int. Acad. Polon. Sci. Cl. Sci. Math. Nat. Ser. A. Sci. Math. (1937) pp. 355-357.

68. L. V. Kantorovič, Functional analysis and applied mathematics (Russian), Uspehi Matem. Nauk (N.S.) vol. 3 (1948) no. 6 pp. 89-185. (A translation is being issued by the National Bureau of Standards, Los Angeles.)

69. L. V. Kantorovič and V. I. Krylov, Приближенные методы высшего анализа (Approximation methods in higher analysis; Russian), 3rd edit., Moscow-Leningrad, 1950, 695 pp.

70. W. Karush, Convergence of a method of solving linear problems, NAML Report 52-19, National Bureau of Standards, Los Angeles, 21 August 1951, multilithed typescript, 20 pp. To appear in Amer. J. Math.

————: see also [52].

70a. M. A. Krasnosel'skiĭ and S. G. Kreĭn, The iterative process with minimal residuals (Russian), Mat. Sbornik (N.S.) vol. 31 (73) (1952) pp. 315-334.

V. I. Krylov: see [69].

71. A. G. Kuroš, A. I. Markuševič, and P. K. Raševskiĭ, Математика в СССР за тридцать лет 1917-1947 (Mathematics in the USSR in the thirty years 1917-1947; Russian), Moscow-Leningrad, 1948, 1044 pp. (Pp. 759-857 on numerical and graphical methods contain a survey and comprehensive bibliography of Russian work.)

72. Cornelius Lanczos, Solution of systems of linear equations by minimized iterations, J. Research Nat. Bur. Standards vol. 49 (1952) pp. 33-53.

73. ————————, Chebyshev polynomials in the solution of large scale linear systems, National Bureau of Standards, Los Angeles, manuscript.

74. H. Liebmann, Die angenäherte Ermittlung harmonischer Funktionen und konformer Abbildung, Sitzungsberichte Bayer. Akad. Wiss. Math.-Nat. Abt. Math.-Phys. Kl. vol. 47 (1918) pp. 385-416.

75. A. T. Lonseth, The propagation of error in linear problems, Trans. Amer. Math. Soc. vol. 62 (1947) pp. 193-212.

76. Samuel Lubkin, A method of summing infinite series, J. Research Nat. Bur. Standards vol. 48 (1952) pp. 228-254.

77. L. A. Lyusternik, Remarks on the numerical solution of boundary problems for Laplace's equation and the calculation of characteristic values by the method of nets (Russian), Trudy Mat. Inst. Steklov vol. 20 (1947) pp. 49-64.

78. C. C. MacDuffee, The theory of matrices, Chelsea, New York, 1946, 110 pp.

79. W. Markoff, Über Polynome, die in einem gegebenen Intervalle möglichst wenig von Null abweichen, Math. Ann. vol. 77 (1916) pp. 213-258. (Translation and condensation by J. Grossman of Russian article published in 1892.)

A. I. Markuševič: see [71].

80. S. N. Mergelyan, Uniform approximation of functions in the complex plane (Russian), Uspehi Matem. Nauk (N.S.) vol. 7 (1952) no. 2, pp. 31-122. Translated in Amer. Math. Soc. Translation T-115.

81. William Edmund Milne, Numerical calculus, Princeton Univ. Press, 1949, 393 pp.

D. Montgomery: see [6].

82. Theodore S. Motzkin, Bibliography on linear inequalities, linear programming, game strategy, economic behavior, and statistical decision functions, in preparation for probable issue by the National Bureau of Standards, Los Angeles.

_____: see also [29, 30, 31, 83].

83. T. S. Motzkin and I. J. Schoenberg, On the relaxation method for linear inequalities, NBS Report 1881, National Bureau of Standards, Los Angeles, August 1952, multilithed typescript, 21 pp. To appear in Proc. Amer. Math. Soc.

84. F. R. Moulton, On the solutions of linear equations having small determinants, Amer. Math. Monthly vol. 20 (1913) pp. 242-249.

85. H. P. Mulholland, On the distribution of a convex even function of several independent rounding-off errors, Proc. Amer. Math. Soc. vol. 3 (1952) pp. 310-321.

86. Francis J. Murray, The theory of mathematical machines, revised edition, King's Crown Press, New York, 1948, 139 pp.

87. National Bureau of Standards, Proceedings of a symposium on simultaneous linear equations and the determination of eigenvalues, August 1951. In press.

88. P. A. Nekrasov, Determination of the unknowns by the method of least squares for very many unknowns (Russian), Mat. Sbornik vol. 12 (1884) pp. 189-204.

89. Alexandre Ostrowski, Sur la détermination des bornes inférieures pour une classe des déterminants, Bull. Sci. Math. Sér. 2, vol. 61 (1937) pp. 19-32.

90. _____, Sur la variation de la matrice inverse d'une matrice donnée, C. R. Acad. Sci. Paris vol. 231 (1950) pp. 1019-1021.

91. _____, Determinants with dominant principal diagonal and the absolute convergence of linear iteration processes (text in German), NBS Report 1727, National Bureau of Standards, Washington, 1 June 1952, multilithed typescript, 49 pp.

92. _____, On the convergence of cyclic linear iterations for symmetric and nearly symmetric matrices. II, NBS Report 1759, National Bureau of Standards, Los Angeles, 26 June 1952, multilithed typescript, 5 pp.

93. _____, On the linear iteration procedures for symmetric matrices, NBS Report 1844, National Bureau of Standards, Los Angeles, 4 August 1952, multilithed typescript, 29 pp.

94. Sam Perlis, Theory of matrices, Addison-Wesley, Cambridge, Mass., 1952, 237 pp.

95. P. Pizzetti, Sulla compensazione delle osservazioni secondo il metodo dei minimi quadrati, nota I, II, Rendiconti della Reale Accademia Nazionale dei Lincei, Rome, Ser. 4, vol. 3 (1887) pp. 230-235 and 288-293.

Hilda Pollaczek-Geiringer: see [44] and [123].

P. K. Raševskiĭ: see [71].

96. Edgar Reich, The solution of linear algebraic equations by successive approximations, Memorandum M-565, Servomechanisms Laboratory, Mass. Inst. of Technology, Cambridge, 5 August 1948, hectographed, 36 pp.

97. _____, On the convergence of the classical iterative method of solving linear simultaneous equations, Ann. Math. Statistics vol. 20 (1949) pp. 448-451.

98. L. F. Richardson, The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam, Philos. Trans. Roy. Soc. London Ser. A. vol. 210 (1910) pp. 307-357.

99. Maria Sofia Roma, Sulla risoluzione numerica dei sistemi di equazioni algebriche lineari col metodo della ortogonalizzazione, Ricerca Sci. vol. 20 (1950) pp. 1288-1290 = Consiglio Naz. Ricerche. Pubbl. Ist. Appl. Calcolo no. 283.

R. L. Rosenberg: see [113].

100. J. Barkley Rosser, Rapidly converging iterative methods for solving linear equations. To appear in [87].

101. Paul A. Samuelson, A convergent iterative process, J. Math. Physics vol. 24 (1945) pp. 131-134.

102. F. E. Satterthwaite, Error control in matrix calculation, Ann. Math. Statistics vol. 15 (1944) pp. 373-387.

I. J. Schoenberg: see [83].

103. E. Schröder, Über unendlich viele Algorithmen zur Auflösung der Gleichungen, Math. Ann. vol. 2 (1870) pp. 317-365.

104. G. Schulz, Iterative Berechnung der reziproken Matrix, Z. Angew. Math. Mech. vol. 13 (1933) pp. 57-59.

105. I. Schur, Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind, J. Reine Angew Math. vol. 147 (1917) pp. 205-232.

106. [Hans Schwerdtfeger], Bibliography on iteration, [reproduced at National Bureau of Standards, Los Angeles, 1951], multilithed typescript, 13 pp. (Bracketed material not on copy.)

107. Ludwig Seidel, Ueber ein Verfahren, die Gleichungen, auf welche die Methode der kleinsten Quadrate führt, sowie lineäre Gleichungen überhaupt, durch successive Annäherung aufzulösen, Abh. Bayer. Akad. Wiss. Math.-Nat. Abt. vol. 11 (1874) no. 3 pp. 81-108.

108. Daniel Shanks, An analogy between transients and mathematical sequences and some nonlinear sequence-to-sequence transforms suggested by it. Part I, NOLM 9994, Naval Ordnance Lab., Silver Spring 19, Md., 26 July 1949, multilithed typescript, 42 pp.

Shortley, George: see [24].

109. R. V. Southwell, Relaxation methods in engineering science, a treatise on approximate computation, Oxford Univ. Press, 1940, 252 pp.

110. _____, Relaxation methods in theoretical physics, Oxford Univ. Press, 1946, 248 pp.

_____: see also [10].

111. J. L. Stearn, Iterative solutions of normal equations, Bull. Géodésique year 1951 (1951) pp. 331-339.

112. Marvin L. Stein, Gradient methods in the solution of systems of linear equations, J. Research Nat. Bur. Standards vol. 48 (1952) pp. 407-413.

_____: see also [53].

113. P. Stein and R. L. Rosenberg, On the solution of linear simultaneous equations by iteration, J. London Math. Soc. vol. 23 (1948) pp. 111-118.

114. E. Stiefel, Über einige Methoden der Relaxationsrechnung, Z. Angew. Math. Physik vol. 3 (1952) pp. 1-33.

_____: see also [54].

W. W. Stifler, Jr.: see [22].

115. O. Taussky, Note on the condition of matrices, Math. Tables and Other Aids to Computation vol. 4 (1950) pp. 111-112.

116. Olga Taussky, Bibliography on bounds for characteristic roots of finite matrices, NBS Report 1162, National Bureau of Standards, Washington, September 1951, multilithed typescript, 10 pp.

117. Olga Taussky and John Todd, Systems of equations, matrices and determinants, Math. Mag. vol. 26 (1952) pp. 9-20 and 00-00.

118. G. Temple, The general theory of relaxation methods applied to linear systems, Proc. Roy. Soc. London Ser. A. vol. 169 (1939) pp. 476-500.

119. John Todd, The condition of a certain matrix, Proc. Cambridge Philos. Soc. vol. 46 (1949) pp. 116-118.

————: see also [117].

120. C. Tompkins, Projection methods in calculation of some linear problems, [1949?], multilithed typescript, 52 pp., part of Engineering Research Associates, Logistics papers, issue no. IV, Appendix I to Bimonthly Progress Report No. 17, Contract N6onr-240, Task Order I, Office of Naval Research, Project NR 047 010.

————: see also [22].

121. L. B. Tuckerman, On the mathematically significant figures in the solution of simultaneous linear equations, Ann. Math. Statistics vol. 12 (1941) pp. 307-316.

122. A. M. Turing, Rounding-off errors in matrix processes, Quart. J. Mech. Appl. Math. vol. 1 (1948) pp. 287-308.

123. R. von Mises and Hilda Pollaczek-Geiringer, Praktische Verfahren der Gleichungsauflösung, Z. Angew. Math. Mech. vol. 9 (1929) pp. 58-77 and 152-164.

J. von Neumann: see [6] and [45].

124. John von Neumann and H. H. Goldstine, Numerical inverting of matrices of high order, Bull. Amer. Math. Soc. vol. 53 (1947) pp. 1021-1099. Part II is [45].

J. H. Wakelin: see [22].

125. Wolfgang Wasow, A note on the inversion of matrices by random walks, Math. Tables and Other Aids to Computation vol. 6 (1952) pp. 78-81.

J. H. Wilkinson: see [35].

L. S. Wilson: see [51].

126. Helmut Wittmeyer, Einfluss der Änderung einer Matrix auf die Lösung des zugehörigen Gleichungssystems, sowie auf die characteristischen Zahlen und die Eigenvektoren, Z. Angew. Math. Mech. vol. 16 (1936), pp. 287-300.

127. _____, Über die Lösung von linearen Gleichungssystemen durch Iteration, Z. Angew. Math. Mech. vol. 16 (1936) pp. 301-310.

128. Max A. Woodbury, Inverting modified matrices, Memorandum Report 42, Statistical Research Group, Princeton, 14 June 1950, hectographed, 4 pp.

129. David M. Young, Jr., Iterative methods for solving partial difference equations of elliptic type, Ph.D. Thesis, Harvard Univ. Math. Dept., 1950, blueprint, c. 100 pp.

130. S. I. Zuhovickiĭ, An algorithm for the solution of the Čebyšev approximation problem in the case of a finite system of incompatable linear equations (Russian), Doklady Akad. Nauk SSSR vol. 79 (1951) pp. 561-564.

131. Rudolf Zurmühl, Matrizen. Eine Darstellung für Ingenieure, Springer, Berlin-Göttingen-Heidelberg, 1950, 427 pp.

National Bureau of Standards, Los Angeles

November 18, 1952

Relaxation

Linear inequalities

General linear Computation

Computing devices

# THE NATIONAL BUREAU OF STANDARDS

## Functions and Activities

The functions of the National Bureau of Standards are set forth in the Act of Congress, March 3, 1901, as amended by Congress in Public Law 619, 1950. These include the development and maintenance of the national standards of measurement and the provision of means and methods for making measurements consistent with these standards; the determination of physical constants and properties of materials; the development of methods and instruments for testing materials, devices, and structures; advisory services to Government Agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; and the development of standard practices, codes, and specifications. The work includes basic and applied research, development, engineering, instrumentation, testing, evaluation, calibration services and various consultation and information services. A major portion of the Bureau's work is performed for other Government Agencies, particularly the Department of Defense and the Atomic Energy Commission. The scope of activities is suggested by the listing of divisions and sections on the inside of the front cover.

## Reports and Publications

The results of the Bureau's work take the form of either actual equipment and devices or published papers and reports. Reports are issued to the sponsoring agency of a particular project or program. Published papers appear either in the Bureau's own series of publications or in the journals of professional and scientific societies. The Bureau itself publishes three monthly periodicals, available from the Government Printing Office: The Journal of Research, which presents complete papers reporting technical investigations; the Technical News Bulletin, which presents summary and preliminary reports on work in progress; and Basic Radio Propagation Predictions, which provides data for determining the best frequencies to use for radio communications throughout the world. There are also five series of nonperiodical publications: The Applied Mathematics Series, Circulars, Handbooks, Building Materials and Structures Reports, and Miscellaneous Publications.

Information on the Bureau's publications can be found in NBS Circular 460, Publications of the National Bureau of Standards ($1.00). Information on calibration services and fees can be found in NBS Circular 483, Testing by the National Bureau of Standards (25 cents). Both are available from the Government Printing Office. Inquiries regarding the Bureau's reports and publications should be addressed to the Office of Scientific Publications, National Bureau of Standards, Washington 25, D. C.


NBS