An Extension of Gauss' Transformation for Improving the Condition of Systems of Linear Equations

1. Gauss' Transformation Extended. Consider a consistent system of linear equations

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \ (i = 1, \cdots, n) \tag{1}$$

with a_{ij}, b_i real. Let the matrix be symmetric and of positive rank $n - d$ and suppose the quadratic form corresponding to A is non-negative semi-definite. Thus the solution points of (1) in affine n-space form a linear subspace of dimension d.

The following is our extension of a transformation due to Gauss: Let $s = (s_1, \cdots, s_n)$ be any real vector. Make the substitution

$$x_i = y_i + s_i y_{n+1} \ (i = 1, \cdots, n), \tag{2}$$

and thereby convert (1) into a system (3) of n equations in the $n + 1$ unknowns $y_1, \cdots, y_{n+1}$:

$$\sum_{j=1}^{n} a_{ij}y_j + \left(\sum_{j=1}^{n} a_{ij}s_j \right) y_{n+1} = b_i \ (i = 1, \cdots, n). \tag{3}$$

An $(n + 1)$-th equation is obtained as the weighted sum of the n equations (3):

$$\sum_{j=1}^{n} \left(\sum_{i=1}^{n} a_{ij}s_i \right) y_j + \left(\sum_{i,j=1}^{n} a_{ij}s_i s_j \right) y_{n+1} = \sum_{i=1}^{n} b_i s_i. \tag{4}$$

The redundancy of (4) means that the solution space of the equation pair (3, 4) is a linear subspace of dimension $d + 1$; that is, the rank of the coefficient matrix A_1 of the system (3, 4) is $n - d$. However, the quantities $x_i = y_i + s_i y_{n+1}$ are determined exactly as well by the system (3, 4) as by the system (1). If A is symmetric, the system (3, 4) also has a symmetric coefficient matrix.

GAUSS[9,10], in writing how he liked to solve certain systems (1) by relaxation,[23] presented a transformation whose application, he was convinced, would improve the convergence of the relaxation process for normal equations associated with the adjustment of surveying data. Gauss' transformation was originally presented only for non-singular ($d = 0$) systems (1), and was the special case $s_1 = \cdots = s_n = -1$ of (2). The same transformation was given by DEDEKIND[5], who showed its effectiveness in one example. ZURMÜHL[22] brings the apparently forgotten transformation to light again, but errs in asserting that it will speed the solution by relaxation and by SEIDEL's method of all (non-singular) systems of equations for which the respective method is slowly convergent.[24]

In two letters Gauss[9,10] reveals the motivation of his transformation $x_i = y_i - y_{n+1}$ in these terms: By the method of least squares he is seeking to determine the values of $n + 1$ quantities $y_1, \cdots, y_{n+1}$ (e.g., azimuths or elevations), whose magnitudes can be deduced from the given data up to

an additive constant. To use the equations (1) amounts to selecting the origin so that $y_{n+1} = 0$ (e.g., measuring angles from one of the unknown azimuths). But how may one decide which unknown to set equal to zero? In this quandary Gauss[10] warns us [p. 251] not to set *any* of the unknowns equal to zero, but to leave them all variable, and then to determine their *differences* by solving the system (3, 4). This, Gauss is convinced, will lead to faster convergence of the relaxation process, because of the symmetrical treatment of all the variables. Incidentally, one also gains an attractive column-sum check as a control on accuracy.

We shall examine the effect of transformation (2) on the system (1) from a different point of view. We shall ascribe a "condition number" $P(A)$ to the matrix A, whether singular or not. We shall show the effect on $P(A)$ of the transformation (2) and, in particular, show when $P(A)$ can be lowered and by how much. As tools we use an extension of a separation lemma known in many connections—for example, for the one-step escalator process for eigenvalues.[14] By repeated application of the extended lemma we derive a k-step separation theorem, believed new, applicable, for example, to the k-step escalator process.[2,8]

For non-singular matrices A, CESARI[4,3] has considered the relation between $P(A)$ and $P[\pi(A)]$, where $\pi(A)$ is a polynomial in A.

For positive definite matrices A the relation of $P(A)$ to the accuracy of the solution of (1) by elimination is discussed at length by VON NEUMANN & GOLDSTINE.[16]

2. Condition of a Singular Matrix. The condition of a system $Ax = b$ with $|A| \neq 0$ describes the influence of small changes in A and b on x; the larger the change in x for given changes in A and b, the "worse" the condition. Though the condition depends also on b, the numbers hitherto proposed (see TODD[18]) to measure the condition are functions solely of A. When A is not singular, Todd suggests the ratio $P = |\lambda_i|_{max}/|\lambda_i|_{min}$ as a condition number of A, where the λ_i are the eigenvalues of A. In the following, however, we are concerned with systems $Ax = b$, where A may be a singular matrix. Then the solutions form a linear subspace X, and it is the displacement of this linear subspace which should be dealt with by a condition number. Cutting with a linear subspace V orthogonal and complementary to X, we can measure the displacement of X by the displacement of its intersection x with V. But x is the unique common point of the intersections of the hyperplanes $Ax = b$ with V. We may therefore measure the condition of the singular system $Ax = b$ by the condition of the related non-singular problem in V. We are thus led to the following definition of a condition number:

Let the eigenvalues λ_i of A be numbered so that

$$(5)\quad 0 = \lambda_1 = \cdots = \lambda_d < |\lambda_{d+1}| \leq |\lambda_{d+2}| \leq \cdots \leq |\lambda_n| \quad (0 \leq d < n).$$

The condition number $P(A)$ of A is defined as the ratio $|\lambda_n|/|\lambda_{d+1}|$ of the maximum and minimum absolute value of the non-vanishing eigenvalues.

For non-negative, semi-definite A all λ_i are real and non-negative. For such A we shall study the effect of the transformation (2) on $P(A)$.

The sensitivity of x or X to changes of the coefficients in (1) probably has a decisive influence on the speed of convergence of an iterative solution of (1). Eigenvalues of A which are exactly zero do not seem to be troublesome in iterative methods of solving the system. In the *gradient method* (see 6, for example) all iterations take place in some subspace V orthogonal to the

solution space X, and one gets to some point of X without difficulty. For the gradient method an immediate extension of theorems of KANTOROVICH[13] proves that the A length[27] of the error vector decreases per step by at least the factor $[P(A) - 1][P(A) + 1]^{-1}$. Perhaps $P(A)$ bears a direct relation to the rate of convergence of iterative processes which are invariant under rotations of the axes. Also, it might ordinarily give some indication of the convergence of processes (like *relaxation* and the methods of Seidel[17] and JACOBI[12]) which are not invariant.

3. Eigenvalues of the Transformed Matrix. To study the effect on P of the transformation (2), we may without loss of generality choose an origin so that each $b_i = 0$ and choose axes so that A is in diagonal form: $a_{ij} = \lambda_i \cdot \delta_{ij}$, where the λ_i are numbered as in (5). Because of the semi-definiteness of A, this can be achieved by a real transformation. The s_i are subjected to the same transformation. Then finding some solution of the d-fold indeterminate system (1) is equivalent to finding some point in the subspace of centers of the family of similar elliptic cylinders

$$\sum_{i=1}^{n} \lambda_i x_i^2 = \text{const.} \tag{6}$$

In the variables y_i defined by (2) the quadrics (6) become a new family of elliptic cylinders. Finding some solution of the $(d + 1)$-fold indeterminate system (3, 4), and hence some solution of (1), is equivalent to finding some point in the subspace of centers of the transformed quadrics

$$Q(y_1, \cdots, y_{n+1}) \equiv \sum_{i=1}^{n} \lambda_i (y_i + s_i y_{n+1})^2 = \text{const.} \tag{7}$$

The geometrical effect of the transformation (2) is easily visualized for a non-singular ($d = 0$) matrix A in two dimensions ($n = 2$). Solving A is equivalent to finding the common center of the ellipses (6). In the variables y_i defined by (2) the ellipses (6) become a family (7) of elliptic cylinders. Each cylinder of (7) is generated by elements parallel to the direction $(s_1, s_2, -1)$ passing through an ellipse (6). Getting a point (y_1, y_2, y_3) on the axis of the cylinders (7) is equivalent to finding one solution of (3, 4). This one solution of (3, 4) yields the unique solution of (1). Note that the eccentricity $\epsilon = \{1 - [P(A_1)]^{-2}\}^{\frac{1}{2}}$ of the elliptical normal sections of the cylinders can be varied in the range $0 \leq \epsilon < 1$ by varying the vector $s = (s_1, s_2)$. The best "condition" of A_1 corresponds to a circular cross section, for which $\epsilon = 0$ and $P(A_1) = 1$.

Let $\mu_0 \leq \mu_1 \leq \cdots \leq \mu_n$ be the eigenvalues of the quadratic form (7), whose matrix A_1 is the coefficient matrix of the system (3, 4). The μ_i are the roots of the determinantal equation

$$\begin{vmatrix} \lambda_1 - \mu & 0 & \cdots & 0 & \lambda_1 s_1 \\ 0 & \lambda_2 - \mu & \cdots & 0 & \lambda_2 s_2 \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & \cdots & \lambda_n - \mu & \lambda_n s_n \\ \lambda_1 s_1 & \lambda_2 s_2 & \cdots & \lambda_n s_n & \sum_{i=1}^{n} \lambda_i s_i^2 - \mu \end{vmatrix} = 0. \tag{8}$$

Expansion of (8) according to the last row and column gives the equation

$$0 = \left(\sum_{i=1}^{n} \lambda_i s_i^2 - \mu\right) \prod_{k=1}^{n} (\lambda_k - \mu) - \sum_{i=1}^{n} \lambda_i^2 s_i^2 \prod_{\substack{k=1 \\ (k \neq i)}}^{n} (\lambda_k - \mu)$$

(9)

$$= -\mu\left[\prod_{k=1}^{n} (\lambda_k - \mu) + \sum_{i=1}^{n} \lambda_i s_i^2 \prod_{\substack{k=1 \\ (k \neq i)}}^{n} (\lambda_k - \mu)\right].$$

The factor μ^{d+1} can be removed from (9), since $\lambda_1 = \cdots = \lambda_d = 0$. There remains the following equation for the other μ_i:

$$\prod_{k=d+1}^{n} (\lambda_k - \mu) + \sum_{i=d+1}^{n} \lambda_i s_i^2 \prod_{\substack{k=d+1 \\ (k \neq i)}}^{n} (\lambda_k - \mu) = 0. \tag{10}$$

We now state the principal tool in the study of (2):

LEMMA. I. *For any real numbers s_i and any set of λ_i satisfying* (5), *the roots μ_i of* (8) *have the following properties: (i) exactly $d + 1$ of the μ_i are zero; (ii) the remaining $n - d$ roots μ_i satisfy the following separation condition:*

$$0 < \lambda_{d+1} \leq \mu_{d+1} \leq \lambda_{d+2} \leq \mu_{d+2} \leq \cdots \leq \lambda_n \leq \mu_n < \infty. \tag{11}$$

II. *Conversely, given any $\mu_{d+1}, \cdots, \mu_n$ satisfying* (11), *one can determine real numbers s_i so that the roots of* (8) *are* 0, 0, $\cdots$, 0, $\mu_{d+1}, \cdots, \mu_n$.

PROOF. *Of* I. Case 1. No $s_i = 0$; $\lambda_{d+1} < \lambda_{d+2} < \cdots < \lambda_n$. We can divide (10) through by $\prod(\lambda_k - \mu)$, getting the equation

$$f(\mu) \equiv \sum_{i=d+1}^{n} \frac{\lambda_i s_i^2}{\mu - \lambda_i} - 1 = 0. \tag{12}$$

Since (12) shows that $f(0) < -1$ and since we previously removed a factor μ^{d+1}, we have proved (*i*). Since each $\lambda_i s_i^2 > 0$, a sketch of $f(\mu)$ shows at once that

$$\lambda_{d+1} < \mu_{d+1} < \lambda_{d+2} < \mu_{d+2} < \cdots < \lambda_n < \mu_n, \tag{13}$$

proving (*ii*).

Case 2. s_i, λ_i unrestricted. Since the roots μ_i of (8) are continuous[25] functions of the λ_i and the s_i, (11) follows from (13) by a passage to the limit.

Of II. Since the choice of $s_1, \cdots, s_d$ is arbitrary, we have only to determine real $s_{d+1}, \cdots, s_n$. This is equivalent to determining non-negative $\lambda_i s_i^2$ $(d + 1 \leq i \leq n)$ so that the roots μ of (10) are the given $\mu_{d+1}, \cdots, \mu_n$.

Case 1. (13) holds. Then equations (10) and (12) are equivalent. But the roots of (12) are the ellipsoidal (confocal) coordinates corresponding to the cartesian coordinates $\{\sqrt{\lambda_i s_i^2}\}$ in $(n - d)$-space. The following inversion formulas give the $\{\lambda_i s_i^2\}$ as rational functions of the μ_j and λ_j; [see[21], p. 548]:

$$\lambda_i s_i^2 = \prod_{j=d+1}^{n} (\mu_j - \lambda_i) \Big/ \prod_{\substack{j=d+1 \\ (j \neq i)}}^{n} (\lambda_j - \lambda_i) > 0 \ (i = d + 1, \cdots, n). \tag{14}$$

Hence $s_{d+1}, \cdots, s_n$ are uniquely determined as positive functions of the λ_j and μ_j, where μ_j are the roots of (10).

Case 2. λ_j, μ_j are restricted only by (11). We shall replace all λ_j, μ_j by neighboring values λ_j', μ_j' which satisfy (13) and also the following condition: for each j where $\lambda_j = \mu_j = \lambda_{j+1}$, we insist that

$$\mu_j' = \tfrac{1}{2}(\lambda_j' + \lambda_{j+1}'). \tag{15}$$

By Case 1, let real $s_i'(d+1 \leq i \leq n)$ be determined so that (14) and (10) hold for the primed symbols. Now let $\lambda_j' \to \lambda_j$, $\mu_j' \to \mu_j$. By (15), the $\lambda_i' s_i'^2$ of (14) all approach (non-negative) limits, which we define to be $\lambda_i s_i^2$. Since the left-hand side of (10) is a continuous function of the arguments λ_j, μ_j, s_j^2, it is seen that (10) is satisfied in the limit. In this manner we have proved the existence of real $s_{d+1}, \cdots, s_n$ in Case 2. (The $s_{d+1}, \cdots, s_n$ need not be unique in Case 2.)

Equation (12) is a special case of the one-step escalator equation of MORRIS.[14] Similar equations occur in the generalized RAYLEIGH-RITZ method of ARONSZAJN[1] (which includes the MORRIS escalator as a special case), in dealing with the realizability of impedance functions by electrical networks[11], and in defining ellipsoidal coordinates.[21] In all these connections Part I of the lemma is known for the case of unequal λ_i.

The lemma may readily be extended to diagonal matrices A with arbitrary real λ_i, although we will not use it. Conclusion (*i*) continues to hold. In addition to condition (11) for the positive λ_i, μ_i, there is a similar condition for the negative λ_i, μ_i, in which, for each i, $\lambda_{i+1} \leq \mu_i \leq \lambda_i < 0$.

4. Effect on the Condition Number. Our condition number for the matrix A is $P(A) = \lambda_n/\lambda_{d+1}$, while the same for the matrix A_1 is $P(A_1) = \mu_n/\mu_{d+1}$. The dependence of $P(A_1)$ on the λ_i and the s_i can ordinarily be stated only in terms of the roots of (8), but certain general remarks can be made:

(a) The lemma shows that $P(A_1)$ can always be made greater than $P(A)$ by some choice of s, and that, unless $\lambda_{d+1} = \lambda_{d+2}$, $P(A_1)$ can also be made less than $P(A)$.

(b) A most favorable choice of s is one for which $\mu_{d+1} = \lambda_{d+2}$ and $\mu_n = \lambda_n$, so that $P(A_1) = \lambda_n/\lambda_{d+2}$. This can be brought about by making (for this particular coordinate system) $s_{d+1} \neq 0$, $s_i = 0$ $(i \neq d+1)$, whence the roots of (8) are 0 $(d+1$ times), $\lambda_{d+2}, \lambda_{d+3}, \cdots, \lambda_n$, and $(s_{d+1}^2 + 1)\lambda_{d+1}$. Then $\mu_{d+1} = \lambda_{d+2}$, $\mu_n = \lambda_n$ if and only if $\lambda_{d+2} \leq (s_{d+1}^2 + 1)\lambda_{d+1} \leq \lambda_n$, or

$$\frac{\lambda_{d+2} - \lambda_{d+1}}{\lambda_{d+1}} \leq s_{d+1}^2 \leq \frac{\lambda_n - \lambda_{d+1}}{\lambda_{d+1}}. \tag{16}$$

In particular, we can choose

$$s_{d+1}^2 = (\lambda_n - \lambda_{d+1})/\lambda_{d+1} = P(A) - 1. \tag{17}$$

(c) For a matrix A not in diagonal form the selection of s such that s_{d+1} satisfies (17) and such that the other $s_i = 0$ can be made as soon as we know λ_{d+1}, λ_n, and the eigenvector u_{d+1} belonging to λ_{d+1}. At least in the usual case $d = 0$ the λ_{d+1}, λ_n, u_{d+1} can ordinarily be approximated by known procedures. To know the least value of $P(A_1)$ achievable by the transformation (2) requires knowledge of λ_{d+2} also. Conversely, if s has been selected so that $\mu_{d+1} = \lambda_{d+2}$, the determination of μ_{d+1}, the least non-zero eigenvalue of A_1,

yields λ_{d+2}. Regarded as a matrix transformation to assist in the determination of the higher eigenvalues of A, this resembles a transformation of TUCKER.[19]

(d) If λ_{d+1}, λ_n, u_{d+1} are known only roughly, we can expect to make $P(A_1)$ reasonably close to its minimum λ_n/λ_{d+2} by picking s in the direction of the rough value of u_{d+1}, with $|s|^2$ equal to the rough value of $(\lambda_n - \lambda_{d+1})/\lambda_{d+1}$.

5. Repeated Application of the Transformation. General Separation Theorem. The transformation (2) can be applied a second time, to generate a matrix A_2 of rank $n - d$ in the $n + 2$ variables $z_1, \cdots, z_{n+2}$. This time 0 becomes a $(d + 2)$-fold multiple eigenvalue of A_2, and the separation formula (11) relates the eigenvalues of A_1 to those of A_2. Finally, the variables z_i and x_i are related by the formula

$$z_i = x_i + s_i z_{n+1} + t_i z_{n+2}. \tag{18}$$

The substitution (18) would border A in one step with two new rows and two new columns. Clearly it is possible for $P(A_2)$ to get as low as λ_n/λ_{d+3}.

If the generalized Gauss transformation (2) is applied k times, we get a matrix A_k of order $n + k$ and rank $n - d$. We have the following theorem, proved by k applications of the lemma:

THEOREM. *The $n + k$ eigenvalues $\kappa_{-k+1}, \cdots, \kappa_0, \kappa_1, \cdots, \kappa_n$ of the matrix A_k have the following properties:* (i) *exactly $d + k$ of the κ_i are zero;* (ii) *the remaining $n - d$ values κ_i can be numbered so as to satisfy the following inequalities:*

$$\begin{cases} \kappa_{d+1} \leq \kappa_{d+2} \leq \cdots \leq \kappa_n; \\ \lambda_i \leq \kappa_i \leq \lambda_{i+k} \ (d + 1 \leq i \leq n - k); \\ \lambda_i \leq \kappa_i < \infty \ (n - k < i \leq n). \end{cases} \tag{19}$$

Conversely, given any $\kappa_{d+1}, \cdots, \kappa_n$ satisfying the inequalities (19), *one can determine k real transformations* (2) *so that A_k, the k-th successive transform of A, has eigenvalues* $0, 0, \cdots, 0, \kappa_{d+1}, \cdots, \kappa_n$.

After k generalized Gaussian transformations (2), we see that $P(A_k)$ can theoretically be made as low as λ_n/λ_{d+k}. After $n - d$ transformations $P(A_{n-d})$ can be made equal to $\lambda_n/\lambda_n = 1$; at this stage the equations are perfectly conditioned.

The theorem can be extended to diagonal matrices A with arbitrary real λ_i. In the extension one gets inequalities of the type

$$\lambda'_{i+k} \leq \kappa_i' \leq \lambda_i'$$

corresponding to negative eigenvalues

$$\lambda'_{i+k} \leq \cdots \leq \lambda'_{i+1} \leq \lambda_i' \leq \cdots < 0$$

of A. In the extended form the theorem is applicable to the k-step ($k > 1$) escalator process of Aronszajn[22] for symmetric matrices A, described by FOX.[8]

6. Example. For a certain class **C** of matrices A it is known *a priori* that $d = 0$ and that u_1 has components which are all positive and roughly equal. **C** includes matrices like (21), which correspond to the Dirichlet problem over a finite net or to related random walk problems. **C** also includes the matrices of the normal equations for the angle variables or the altitudes

in a survey; this was the source of the examples in[9, 10, 5]. If A belongs to **C**, the original form [$s = (-1, \cdots, -1)$ in unnormalized coordinates] of Gauss' transformation is close to the optimal selection of s parallel to u_1.

On the other hand, if A is such that $d = 0$ but $(-1, \cdots, -1)$ is, roughly speaking, closer to u_n than to u_1, Gauss' form of (2) is likely to make $P(A_1) > P(A)$, whereas a choice of s near u_1 will make $P(A_1) \leq P(A)$.

As an example of this, we cite T. S. WILSON's ill-conditioned matrix

$$A = \begin{bmatrix} 5 & 7 & 6 & 5 \\ 7 & 10 & 8 & 7 \\ 6 & 8 & 10 & 9 \\ 5 & 7 & 9 & 10 \end{bmatrix} \tag{20}$$

given by Todd.[18] The decomposition of A is as follows:[26]

$$\begin{aligned} \lambda_1 &= 0.01015, & u_1 &= (-.830, \ .501, \ .208, -.124); \\ \lambda_2 &= 0.8431, & u_2 &= (\ .094, -.302, \ .761, -.568); \\ \lambda_3 &= 3.858, & u_3 &= (\ .396, \ .614, -.271, -.625); \\ \lambda_4 &= 30.29, & u_4 &= (\ .380, \ .526, \ .552, \ .521). \end{aligned}$$

Hence $P(A) = 2984$, a relatively high value.

Gauss' original transformation, corresponding to $s = (-1, -1, -1, -1)$ in unnormalized coordinates, would replace A by the coefficient matrix

$$A_1 = \begin{bmatrix} 5 & 7 & 6 & 5 & -23 \\ 7 & 10 & 8 & 7 & -32 \\ 6 & 8 & 10 & 9 & -33 \\ 5 & 7 & 9 & 10 & -31 \\ -23 & -32 & -33 & -31 & 119 \end{bmatrix}.$$

To obtain the condition number of A_1, we find the components s_i of s in the coordinate system of the eigenvectors u_i:

$$s_1 = .245, \ s_2 = .015, \ s_3 = -.114, \ s_4 = -1.979.$$

With these s_i, equation (12) becomes

$$\frac{.00061}{\mu - .01015} + \frac{.0002}{\mu - .8431} + \frac{.050}{\mu - 3.858} + \frac{118.6}{\mu - 30.29} = 1.$$

Hence the eigenvalues of A_1 are approximately

$$\mu_1 = .01027, \ \mu_2 = .8431, \ \mu_3 = 3.867, \ \mu_4 = 148.9,$$

so that $P(A_1) \doteq 14500$. As measured by P, the matrix A_1 resulting from Gauss' original transformation is even worse conditioned than A.

On the other hand, some rough knowledge of λ_1, λ_4 permits considerable reduction in P. Following the principles of section 4 but using only one figure of u_1, we select s in the direction $(.8, -.5, -.2, .1)$. To satisfy (17) approximately we multiply this vector by 50, getting $s = (40, -25, -10, 5)$. With these weights we obtain the transformed matrix

$$A_1' = \begin{bmatrix} 5 & 7 & 6 & 5 & -10 \\ 7 & 10 & 8 & 7 & -15 \\ 6 & 8 & 10 & 9 & -15 \\ 5 & 7 & 9 & 10 & -15 \\ -10 & -15 & -15 & -15 & 50 \end{bmatrix}.$$

The components of s in the normalized coordinate system are

$$s_1 = -48.42,\ s_2 = .86,\ s_3 = .075,\ s_4 = -.865.$$

With these s_i equation (12) becomes

$$\frac{23.797}{\mu - .01015} + \frac{.623}{\mu - .8431} + \frac{.0217}{\mu - 3.858} + \frac{22.66}{\mu - 30.29} = 1.$$

Hence the eigenvalues of A_1' are approximately

$$\mu_1 = .8205,\ \mu_2 = 3.853,\ \mu_3 = 11.21,\ \mu_4 = 66.22,$$

so that $P(A_1') \doteq 80.71$. Thus A_1' is far better conditioned than A. If we had selected s *exactly* parallel to u_1, $P(A_1')$ could have been made as low as $(30.29)/(.8431) = 35.93$.

7. Pairing of the Eigenvalues. By the lemma it is theoretically always possible for n even, $d = 0$, to make the non-zero eigenvalues μ_i occur in pairs, even though all λ_i are distinct; cf.[20]. If so, in solving (3, 4) by the gradient method, for example, the double roots μ_i of A_1 act like single roots and the essential dimensionality of the calculation is reduced from n to $n/2$. For example, consider

$$(21)\quad A = \begin{bmatrix} 2 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{bmatrix} (2m \text{ rows}),$$

whose eigenvalues are $4 \sin^2 [k\pi/(2m + 1)]$ $(k = 1, 2, \cdots, 2m)$. If $s = (-1, \cdots, -1)$, the transformation (2) yields the matrix

$$(22)\quad A_1 = \begin{bmatrix} 2 & -1 & 0 & 0 & \cdots & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ -1 & 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{bmatrix} (2m + 1 \text{ rows}),$$

whose eigenvalues are $4 \sin^2 [2k\pi/(2m + 1)]$ $(k = 0, 1, 1, 2, 2, \cdots, m, m)$.

These two matrices are related to those for discrete random walks in one dimension: (21) to walks on a line segment with "manholes" at both ends, and (22) to walks on the circumference of a circle, with no manholes. The matrices are also used in solving the Dirichlet problem on a discrete net.

NBSINA
Univ. of Calif.
Los Angeles

GEORGE E. FORSYTHE
THEODORE S. MOTZKIN

The preparation of this paper was sponsored in part by the Office of Scientific Research, U. S. Air Force, and by the Office of Naval Research.

[1] N. ARONSZAJN, "Rayleigh-Ritz and A. Weinstein methods for approximation of eigenvalues. I. Operators in a Hilbert space. II. Differential operators," Nat. Acad. Sci., *Proc.*, v. 34, 1948, p. 474–480 and p. 594–601.

[2] N. ARONSZAJN, "Escalator and modified escalator methods" (unpublished manuscript, 1945, 16 pp.).

[3] E. BODEWIG, "Bericht über die verschiedenen Methoden zur Lösung eines Systems linearen Gleichungen mit reellen Koeffizienten. III," Akad. Wetensch., Amsterdam, *Proc.*, v. 50, 1947, p. 1285–1295 = *Indagationes Math.*, v. 9, 1947, p. 611–621.

[4] LAMBERTO CESARI, "Sulla risoluzioni dei sistemi di equazioni lineari per approssimazioni successive," Reale Accad. dei Lincei, *Classe scienze fis., mat., natur., Rendic.*, v. 25, s. 6a, 1937, p. 422–428.

[5] R. DEDEKIND, "Gauss in seiner Vorlesung über die Methode der kleinsten Quadrate," *Festschrift zur Feier des 150-jährigen Bestehen der königlichen Gesellschaft der Wissenschaften zu Göttingen.* Berlin, 1901, p. 45–59.

[6] GEORGE E. FORSYTHE & THEODORE S. MOTZKIN, "Acceleration of the optimum gradient method. Preliminary report," [Abstract] Amer. Math. Soc., *Bull.*, v. 57, 1951, p. 304–305.

[7] L. FOX, "A short account of relaxation methods," *Quart. Jn. Mech. Appl. Math.*, v. 1, 1948, p. 253–280.

[8] L. Fox, "Escalator methods for latent roots," *Quart. Jn. Mech. Appl. Math.*, to appear in v. 5, 1952.

[9] C. F. GAUSS, "Letter to Gerling, 26 December 1823," *Werke* v. 9, p. 278–281. For an annotated translation of Gauss' letter by G. E. Forsythe, see *MTAC*, v. 5, p. 155–258.

[10] C. F. GAUSS, "Letter to Gerling, 19 January 1840," *Werke* v. 9, p. 250–253.

[11] ERNST A. GUILLEMIN, *Communications Networks.* V. 2, New York, 1935, p. 187.

[12] C. G. J. JACOBI, "Ueber eine neue Auflösungsart der bei der Methode der kleinsten Quadrate vorkommenden lineären Gleichungen," *Astronomische Nachrichten*, v. 22, 1845, no. 523, cols. 297–306.

[13] L. V. KANTOROVICH, "Functional analysis and applied mathematics," *Uspekhi Matematicheskikh Nauk*, n. s., v. 3, no. 6 (28), 1948, p. 89–185. [Russian].

[14] JOSEPH MORRIS, *The Escalator Method*, New York, 1947, p. 111–112.

[15] THEODORE MOTZKIN, "From among n conjugate algebraic integers, $n-1$ can be approximately given," Amer. Math. Soc., *Bull.*, v. 53, 1947, p. 156–162.

[16] JOHN VON NEUMANN & H. H. GOLDSTINE, "Numerical inverting of matrices of high order," Amer. Math. Soc., *Bull.*, v. 53, 1947, p. 1021–1099, and Amer. Math. Soc., *Proc.*, v. 2, 1951, p. 188–202.

[17] LUDWIG SEIDEL, "Ueber ein Verfahren, die Gleichungen, auf welche die Methode der kleinsten Quadrate führt, sowie lineäre Gleichungen überhaupt, durch successive Annäherung aufzulösen," Akad. Wiss., Munich, *mat.-nat. Abt., Abhandlungen*, v. 11, no. 3, 1874, p. 81–108.

[18] JOHN TODD, "The condition of a certain matrix," Camb. Phil. Soc., *Proc.*, v. 46, 1949, p. 116–118.

[19] L. R. TUCKER, "The determination of successive principal components without computation of tables of residual correlation coefficients," *Psychometrika*, v. 9, 1944, p. 149–153.

[20] ALEXANDER WEINSTEIN, "Separation theorems for the eigenvalues of partial differential equations," Reissner Anniversary Volume, *Contributions to Applied Mechanics.* Ann Arbor, 1949.

[21] E. T. WHITTAKER & G. N. WATSON, *A Course of Modern Analysis.* American edition, New York, 1943, p. 547.

[22] R. ZURMÜHL, *Matrizen. Eine Darstellung für Ingenieure.* Berlin, 1949, p. 280–282.

[23] Gauss was very fond of relaxation, which is identical with the process summarized by Fox.[7] Gauss[9] remarked that the process was so easy that he could do it while half asleep or while thinking about other things.

[24] Consider the system $x + ry = 0$, $rx + y = 0$, $|r| < 1$. When $r = \pm(1 - \epsilon)$, either method converges slowly. Gauss' transformation with $s = (-1, -1)$ very much speeds the relaxation solution and the Seidel solution for $r = -(1 - \epsilon)$; but for $r = +(1 - \epsilon)$ it does not affect the relaxation solution and worsens the Seidel solution.

[25] For a simple proof of this well known fact, see Motzkin.[15]

[26] The calculations of this section were performed by Mrs. LOUISE STRAUS.

[27] The A length of a vector u is $(\Sigma a_{ij}u_iu_j)^{\frac{1}{2}}$.