

On Lanczos' Algorithm for Tri-Diagonalization

Tetsuro YAMAMOTO*

(Received September 17, 1968)

The Lanczos algorithm transforming a given matrix into a tri-diagonal form is well known in numerical analysis and is discussed in many literatures. The possibility of this algorithm is shown in Rutishauser's excellent paper [8]. However it seems to the author that no further theoretical consideration has been made since then.

The process starts from a pair of trial vectors x_1 and y_1 . A pair of the i -th iterated vectors x_i and y_i can be constructed successively if $y_j^* x_j \neq 0$ ($1 \leq j \leq i-1$). Hence, if $y_{p+1}^* x_{p+1} = 0$ for some $p \leq n-1$, we must modify the algorithm so as to continue. This is possible in case where $x_{p+1} = 0$ or $y_{p+1} = 0$, while any method of modification is not known in case where $x_{p+1} \neq 0$ and $y_{p+1} \neq 0$. We shall call the former case "lucky" and the latter "unlucky". The only thing for us to do in "unlucky" case is to choose new starting vectors x_1, y_1 and begin again in the hope that this case will not happen later. Rutishauser's result ([8] Satz 1) guarantees this possibility.

In practical computation, however, "unlucky" case may occur after repeated modifications in "lucky" cases. Once we encountered with "unlucky" case, we have to abandon all the efforts made before and start again with new trial vectors (if we stick to the old knowledge). Then a question arises naturally: Is it actually necessary to go back to the first step? In this paper we shall treat this problem. Roughly speaking, the answer is as follows: It is sufficient to go back to the latest modification. As a special case of this result, we can show that one of the initial vectors can be chosen arbitrarily to avoid "unlucky" case. Further it will be shown that there exists a vector x such that the algorithm starting from $x_1 = y_1 = x$ can be continued so that "unlucky" case may not occur. These results will be stated in Theorems 3-6 of §2 and a new procedure will be proposed at p. 279. Finally, in connection with the Lanczos algorithm, we shall give, in Appendix, some properties concerning the location of the eigenvalues of tri-diagonal matrices.

§1. Preliminaries

1.1. Let A be a given (complex or real) matrix of order n . Starting from a pair of initial vectors x_1 and y_1 , construct a sequence of iterated vectors x_i, y_i as follows:

* This work was partially supported by a research grant of the Sakkokai Foundation.

Case 3. $x_{p+1} \neq 0, \quad y_{p+1} = 0,$

Case 4. $x_{p+1} \neq 0, \quad y_{p+1} \neq 0.$

For the first three cases, we can continue the process by the following modification:

Case 1. In this case, take a non-zero vector w_{p+1} which is orthogonal to the vectors x_1, x_2, \dots, x_p . Then there exists a vector z_{p+1} orthogonal to the vectors y_1, y_2, \dots, y_p such that $w_{p+1}^* z_{p+1} \neq 0$. In fact the whole space (n -dimensional Euclidean or unitary space) is a direct sum of the space spanned by x_1, x_2, \dots, x_p and the orthogonal complement of the space spanned by y_1, y_2, \dots, y_p . Thus, if we replace x_{p+1} by z_{p+1} and y_{p+1} by w_{p+1} , then we can continue the process by the formulas¹⁾

$$\begin{aligned} x_{p+2} &= Az_{p+1} - \alpha'_{p+1} z_{p+1} \\ y_{p+2} &= A^* w_{p+1} - \bar{\alpha}'_{p+1} w_{p+1}, \quad \alpha'_{p+1} = \frac{w_{p+1}^* A z_{p+1}}{w_{p+1}^* z_{p+1}} \\ x_{p+3} &= Ax_{p+2} - \alpha_{p+2} x_{p+2} - \beta'_{p+1} z_{p+1}, \\ y_{p+3} &= A^* y_{p+2} - \bar{\alpha}_{p+2} y_{p+2} - \bar{\beta}'_{p+1} w_{p+1}, \quad \beta'_{p+1} = \frac{y_{p+2}^* x_{p+2}}{w_{p+1}^* z_{p+1}}, \\ x_{p+i} &= Ax_{p+i-1} - \alpha_{p+i-1} x_{p+i-1} - \beta_{p+i-2} x_{p+i-2}, \\ y_{p+i} &= A^* y_{p+i-1} - \bar{\alpha}_{p+i-1} y_{p+i-1} - \bar{\beta}_{p+i-2} y_{p+i-2} \quad (i \geq 4). \end{aligned}$$

Case 2. In this case, by similar argument, we can prove the existence of a vector z_{p+1} such that $y_j^* z_{p+1} = 0 \quad (1 \leq j \leq p)$ and $y_{p+1}^* z_{p+1} \neq 0$. Therefore the modified formulas in this case are

$$\begin{aligned} x_{p+2} &= Az_{p+1} - \alpha'_{p+1} z_{p+1} - \beta'_p x_p, \\ \alpha'_{p+1} &= \frac{y_{p+1}^* A z_{p+1}}{y_{p+1}^* z_{p+1}}, \quad \beta'_p = \frac{y_{p+1}^* z_{p+1}}{y_p^* x_p}, \\ y_{p+2} &= A^* y_{p+1} - \bar{\alpha}'_{p+1} y_{p+1}, \\ x_{p+3} &= Ax_{p+2} - \alpha_{p+2} x_{p+2} - \beta'_{p+1} z_{p+1}, \\ y_{p+3} &= A^* y_{p+2} - \bar{\alpha}_{p+2} y_{p+2} - \bar{\beta}'_{p+1} y_{p+1}, \quad \beta'_{p+1} = \frac{y_{p+2}^* x_{p+2}}{y_{p+1}^* z_{p+1}}, \\ x_{p+i} &= Ax_{p+i-1} - \alpha_{p+i-1} x_{p+i-1} - \beta_{p+i-2} x_{p+i-2}, \\ y_{p+i} &= A^* y_{p+i-1} - \bar{\alpha}_{p+i-1} y_{p+i-1} - \bar{\beta}_{p+i-2} y_{p+i-2} \quad (i \geq 4). \end{aligned}$$

1) For detailed discussion, see [3].

Case 3. This case is similar to Case 2.

For Case 4, however, any method of such modification is not known. The only thing to do in this case is to choose new starting vectors x_1 and y_1 and begin again in the hope that this case will not happen later.

1.3. We note that, even if we modify the procedure for Cases 1-3 so as to continue, Case 4 may occur at the later step. For example, let

$$A = \begin{pmatrix} 5/9 & -2/9 & 4/9 & 0 \\ 4/9 & 2/9 & 5/9 & 0 \\ -2/9 & 8/9 & 2/9 & 0 \\ 1/3 & 2/3 & 2/3 & 1 \end{pmatrix}$$

and choose a pair of initial vectors $x_1 = {}^t(2/3, 1/3, -2/3, 0)$ and $y_1 = {}^t(1/3, 1/6, -1/3, 1)$. Then, by simple computation we have $x_2 = 0$ and $y_2 = {}^t(2/3, 1/3, 5/6, 1)$; namely Case 2 occurs. Hence, according to Causey and Gregory's proposal [3], take a new vector

$$z_2 = y_2 - \frac{y_1^* y_2}{y_1^* x_1} x_1 = {}^t(-2/3, -1/3, 13/6, 1)$$

which satisfies $y_2^* z_2 \neq 0$ and $y_1^* z_2 = 0$. Then Case 4 will occur and the algorithm fails there. In fact we have

$$x_3 = {}^t(-13/9, -2/9, 4/9, 2/3),$$

$$y_3 = {}^t(-2/9, 8/9, 2/9, -1/3),$$

and

$$y_3^* x_3 = 0.$$

Therefore, if we obey the old principle, we shall have to go back to the first step and start again from new vectors x_1 and y_1 . However, this is not only inefficient, but also unnecessary. For, as is easily seen, if we take another vector $z'_2 = {}^t(-1/3, -2/3, 7/3, 1)$ and start again from there with a pair of vectors z'_2 and y_2 , the algorithm²⁾ can be well continued to completion.³⁾ The above fact is true in general. The purpose of this paper is to show this and give an improved procedure for the Lanczos algorithm.

1.4. Notations and definitions. Throughout this paper, we consider

2) In the following, a term "the (Lanczos) algorithm" stands for the procedure according to the modification mentioned above when Cases 1-3 occurred.

3) Namely, in the sense of footnote 2), the algorithm can be continued so that Case 4 may not occur. In the following we shall often use this expression.

complex (or real) matrices and certain notational conventions will be observed. For a matrix A , A^* (tA) stands for a conjugate transposed (transposed) matrix of A . $A_{(j_1, j_2, \dots, j_r)}^{(i_1, i_2, \dots, i_r)}$ denotes an r -square submatrix obtained from the i_1, i_2, \dots, i_r th rows and j_1, j_2, \dots, j_r th columns. A square matrix A is called non-derogatory if its minimal polynomial is the same as its characteristic polynomial, and otherwise called derogatory. A square matrix $A=(a_{ij})$ is of an upper Hessenberg type if $a_{ij}=0$ for $i-j \geq 2$. For an n -square matrix A and an n -dimensional vector x , there exists a number $p(\leq n)$ such that a set of vectors $x, Ax, \dots, A^{p-1}x$ is linearly independent and a set of vectors $x, Ax, \dots, A^{p-1}x, A^p x$ is linearly dependent. The number p is called the grade of x with respect to A .⁴⁾ Clearly x is a vector of grade p with respect to A if and only if $\varphi(A)x=0$ for a unique monic polynomial $\varphi(\lambda)$ of degree p and $\psi(A)x \neq 0$ for any polynomial $\psi(\lambda)$ of degree less than p . Let x_1, x_2, \dots, x_m be a set of m vectors. Then we denote by $[x_1, x_2, \dots, x_m]$ and $[x_1, x_2, \dots, x_m]^\perp$ the vector subspace spanned by x_1, x_2, \dots, x_m and the orthogonal complement of the subspace $[x_1, x_2, \dots, x_m]$ respectively. Finally, for vectors $a_1 = {}^t(a_{11}, a_{12}, \dots, a_{1p})$, $a_2 = {}^t(a_{21}, a_{22}, \dots, a_{2q})$, \dots , $a_s = {}^t(a_{s1}, a_{s2}, \dots, a_{sr})$, the notation $a_1 \oplus a_2 \oplus \dots \oplus a_s$ or $\sum_{i=1}^s \oplus a_i$ means a vector ${}^t(a_{11}, a_{12}, \dots, a_{1p}, a_{21}, a_{22}, \dots, a_{2q}, \dots, a_{s1}, a_{s2}, \dots, a_{sr})$. Similarly, for submatrices A_1, A_2, \dots, A_s , we shall use the notation $A_1 \oplus A_2 \oplus \dots \oplus A_s$ or $\sum_{i=1}^s \oplus A_i$ in place of a matrix

$$\left(\begin{array}{c} A_1 \\ A_2 \\ \dots \\ A_s \end{array} \right).$$

§2. The possibility of Lanczos' algorithm

2.1. We begin with

THEOREM 1. *Let A be a given matrix of order n . Then, by the Lanczos algorithm starting from appropriate vectors x_1 and y_1 , we can always get a Jordan normal form.*

PROOF. Take a non-singular matrix T such that $T^{-1}AT$ is a Jordan normal form; i.e., $T^{-1}AT = \sum_{i=1}^s \oplus J_i$ where J_i are of order n_i and

$$J_i = \left(\begin{array}{ccc} \lambda_i & & \\ 1 & \lambda_i & \\ & \dots & \dots \\ & & 1 & \lambda_i \end{array} \right).$$

4) By definition, it is clear that the grade of any vector with respect to A does not exceed the degree of its minimal polynomial. Hence we note here that, if A is derogatory, the breakdown of the algorithm (i.e., Cases 1-4) will certainly occur.

Let t_i and u_i be the i -th columns of T and T^{*-1} respectively. Then, by the Lanczos algorithm starting from $x_1=t_1$ and $y_1=u_1$, we get $x_2=t_2$ and $y_2=0$. Thus, according to the modification mentioned in §1, we can choose u_2 as the vector w_2 . Repeating the similar modification, the algorithm can be continued to completion as follows:

$$\begin{aligned} x_i &= t_i & (i \neq n_1+1, n_1+n_2+1, \dots, n_1+\dots+n_{s-1}+1), \\ x_{n_1+\dots+n_i+1} &= 0, & z_{n_1+\dots+n_i+1} = t_{n_1+\dots+n_i+1} \quad (1 \leq i \leq s-1), \\ y_i &= 0, & w_i = u_i \quad (2 \leq i \leq n), \end{aligned}$$

and

$$\beta_i = 0 \quad (1 \leq i \leq n-1).$$

Hence the result is

$$\begin{aligned} &A(x_1, \dots, x_{n_1}, z_{n_1+1}, x_{n_1+2}, \dots, x_{n_1+n_2}, z_{n_1+n_2+1}, \dots, x_n) \\ &= (x_1, \dots, x_{n_1}, z_{n_1+1}, x_{n_1+2}, \dots, x_n) \cdot (J_1 \oplus J_2 \oplus \dots \oplus J_s). \end{aligned}$$

This proves Theorem 1.

The above proof shows that theoretically a Jordan normal form can be obtained by executing the Lanczos algorithm, using only the modification for Case 1. If A is a real matrix and all the eigenvalues of A are real, then T , or t_i and u_i , may be taken to be real. Therefore, in such a case we can obtain a Jordan normal form by using the algorithm in the realm of real numbers. In practical computation, however, it is difficult to find the initial vectors t_1, u_1 , etc. Hence we are to seek for other properties which assure the possibility of the algorithm.

2.2. The following lemma plays a fundamental role throughout this paper.

LEMMA 1. *Let \tilde{A} and A be matrices such that $\tilde{A} = T^{-1}AT$ with a non-singular matrix T . If we denote by \tilde{x}_i, \tilde{y}_i (x_i, y_i) the iterated vectors obtained by Lanczos' algorithm for \tilde{A} (A) with initial vectors \tilde{x}_1, \tilde{y}_1 ($x_1 = T\tilde{x}_1, y_1 = T^{*-1}\tilde{y}_1$), then we have $x_i = T\tilde{x}_i, y_i = T^{*-1}\tilde{y}_i$. Hence $x_{p+1} = 0$ ($y_{p+1} = 0$) for some p if and only if $\tilde{x}_{p+1} = 0$ ($\tilde{y}_{p+1} = 0$). Further, if we take a modified vector \tilde{z}_{p+1} ⁵⁾ (\tilde{w}_{p+1}) for \tilde{A} , then $z_{p+1} = T\tilde{z}_{p+1}$ ($w_{p+1} = T^{*-1}\tilde{w}_{p+1}$) is a modified vector for A .*

PROOF. The following relations hold:

$$\tilde{x}_{i+1} = (T^{-1}AT)\tilde{x}_i - \frac{\tilde{y}_i^* T^{-1}AT\tilde{x}_i}{\tilde{y}_i^* \tilde{x}_i} \tilde{x}_i - \frac{\tilde{y}_i^* \tilde{x}_i}{\tilde{y}_{i-1}^* \tilde{x}_{i-1}} \tilde{x}_{i-1},$$

and

5) Namely, $\tilde{z}_{p+1} \in [\tilde{y}_1, \dots, \tilde{y}_p]^\perp$ and $\tilde{y}_{p+1}^* \tilde{z}_{p+1} \neq 0$ (if $\tilde{y}_{p+1} \neq 0$), etc.

$$\tilde{y}_{i+1} = T^* A^* T^{*-1} \tilde{y}_i - \frac{\tilde{x}_i^* T^* A^* T^{*-1} \tilde{y}_i}{\tilde{x}_i^* \tilde{y}_i} \tilde{y}_i - \frac{\tilde{x}_i^* \tilde{y}_i}{\tilde{x}_{i-1}^* \tilde{y}_{i-1}} \tilde{y}_{i-1},$$

which may be written as

$$T\tilde{x}_{i+1} = A(T\tilde{x}_i) - \frac{(T^{*-1}\tilde{y}_i)^* A(T\tilde{x}_i)}{(T^{*-1}\tilde{y}_i)^*(T\tilde{x}_i)} T\tilde{x}_i - \frac{(T^{*-1}\tilde{y}_i)^*(T\tilde{x}_i)}{(T^{*-1}\tilde{y}_{i-1})^*(T\tilde{x}_{i-1})} T\tilde{x}_{i-1},$$

and

$$T^{*-1}\tilde{y}_{i+1} = A^*(T^{*-1}\tilde{y}_i) - \frac{(T\tilde{x}_i)^* A^*(T^{*-1}\tilde{y}_i)}{(T\tilde{x}_i)^*(T^{*-1}\tilde{y}_i)} T^{*-1}\tilde{y}_i - \frac{(T\tilde{x}_i)^*(T^{*-1}\tilde{y}_i)}{(T\tilde{x}_{i-1})^*(T^{*-1}\tilde{y}_{i-1})} T^{*-1}\tilde{y}_{i-1}.$$

This implies that $T\tilde{x}_i$ and $T^{*-1}\tilde{y}_i$ is the i -th iterated vectors for A starting from initial vectors $T\tilde{x}_1$ and $T^{*-1}\tilde{y}_1$. Similarly the remaining part can be verified easily. Q.E.D.

2.3. The following Lemmas 2, 3 and Theorem 2 are due to Rutishauser [8]. But we give here purely algebraic proofs of Lemma 2 and Theorem 2 for the sake of completeness.

LEMMA 2. *Let A be a matrix of order n and m be the degree of the minimal polynomial for A . If we put*

$$f_i(x, y, A) = \begin{vmatrix} y^*x & y^*Ax & y^*A^{i-1}x \\ y^*Ax & y^*A^2x & \dots & y^*A^i x \\ \dots & \dots & \dots & \dots \\ y^*A^{i-1}x & y^*A^i x & \dots & y^*A^{2i-2}x \end{vmatrix}$$

with n -dimensional vectors x and y , then there exist two vectors x_1 and y_1 such that $f_i(x_1, y_1, A) \neq 0$ ($1 \leq i \leq m$).

PROOF. Let T be a non-singular matrix such that $T^{-1}AT = A_1 \oplus A_2$, where A_1 is of order m , A_2 of order $n - m$, and

$$A_1 = \left(\begin{array}{c|c} \begin{matrix} \lambda_1 & & & \\ 1 & \lambda_1 & & \\ & \ddots & \ddots & \\ & & 1 & \lambda_1 \end{matrix} & \\ \hline & \begin{matrix} \lambda_2 & & & \\ 1 & \lambda_2 & & \\ & \ddots & \ddots & \\ & & 1 & \lambda_2 \end{matrix} \\ \hline & & & \ddots \end{array} \right) \quad (\lambda_i \neq \lambda_j \text{ for } i \neq j).$$

$$(3) \quad L_i = \begin{pmatrix} \alpha_{i1} & \beta_{i1} & & & \\ & 1 & \alpha_{i2} & \beta_{i2} & \\ & & 1 & \ddots & \\ & & & \ddots & \beta_{in_i-1} \\ & & & & 1 & \alpha_{in_i} \end{pmatrix} \quad (\beta_{ij} \neq 0 \text{ for any } i, j)$$

and $m = n_1 \geq n_2 \geq \dots \geq n_s$.⁶⁾

PROOF. Considering a Jordan normal form, there exists a non-singular matrix T such that $T^{-1}AT = \sum_{i=1}^s \oplus A_i$, where A_i are non-derogatory of order n_i ($m = n_1 \geq n_2 \geq \dots \geq n_s$, $\sum_{i=1}^s n_i = n$) and the characteristic polynomial for A_i coincides with the minimal polynomial of $\sum_{j=i}^s \oplus A_j$ ($1 \leq i \leq s$). Then, by Lemma 3, we can find two n_i -dimensional vectors x_{i1}, y_{i1} such that $y_{i1}^* x_{i1} \neq 0$ and the j -th iterated vectors x_{ij}, y_{ij} for A_i starting from x_{i1}, y_{i1} satisfy

$$y_{ik}^* x_{ij} \begin{cases} = 0 & (j \neq k) \\ \neq 0 & (j = k) \end{cases} \quad (1 \leq j, k \leq n_i),$$

and

$$x_{in_i+1} = y_{in_i+1} = 0$$

for each i ($1 \leq i \leq s$). Therefore, by Lemma 1, if we apply the algorithm to A with initial vectors

$$x_1 = T \cdot {}^t(x_{11}, \overbrace{0, \dots, 0}^{n-n_1}), \quad y_1 = T^{*-1} \cdot {}^t(y_{11}, \overbrace{0, \dots, 0}^{n-n_1}),$$

the iterated vectors x_j, y_j must have the form

$$x_j = T \cdot {}^t(x_{1j}, 0, \dots, 0), \quad y_j = T^{*-1} \cdot {}^t(y_{1j}, 0, \dots, 0) \quad (1 \leq j \leq n_1)$$

and

$$x_{n_1+1} = y_{n_1+1} = 0.$$

Next we set

$$z_{n_1+1} = T \cdot {}^t(\overbrace{0, \dots, 0}^{n_1}, {}^t x_{21}, \overbrace{0, \dots, 0}^{n-n_1-n_2}),$$

$$w_{n_1+1} = T^{*-1} \cdot {}^t(0, \dots, 0, {}^t y_{21}, 0, \dots, 0)$$

and begin again with them, since

$$z_{n_1+1} \in [y_1, \dots, y_{n_1}]^\perp, \quad w_{n_1+1} \in [x_1, \dots, x_{n_1}]^\perp,$$

6) More precise results will be given later as Theorems 5 and 6.

and

$$w_{n_1+1}^* z_{n_1+1} \neq 0.$$

Then the iterated vectors x_j, y_j satisfy

$$y_j^* x_j \neq 0 \quad (n_1 + 1 \leq j \leq n_1 + n_2)$$

and

$$x_{n_1+n_2+1} = y_{n_1+n_2+1} = 0.$$

Continuing this process, the algorithm is complete after $s-1$ modifications in Case 1 and the result is

$$AX = X(L_1 \oplus L_2 \oplus \cdots \oplus L_s)$$

where

$$X = (x_1, \dots, x_{n_1}, z_{n_1+1}, x_{n_1+2}, \dots, x_{n_1+\dots+n_{s-1}}, z_{n_1+\dots+n_{s-1}+1}, \dots, x_n),$$

$$z_{n_1+\dots+n_i+1} = T \cdot \begin{pmatrix} 0, \dots, 0, & {}^t x_{i+1}, & 0, \dots, 0 \end{pmatrix},$$

$$\underbrace{\hspace{10em}}_{n_1+\dots+n_i} \quad \underbrace{\hspace{10em}}_{n-n_1-\dots-n_{i+1}}$$

$$x_{n_1+\dots+n_i+j} = T \cdot \begin{pmatrix} 0, \dots, 0, & {}^t x_{i+1}, & 0, \dots, 0 \end{pmatrix} \quad (2 \leq j \leq n_{i+1}),$$

and L_i is a non-derogatory tri-diagonal matrix shown in (3) with

$$\alpha_{ij} = \frac{y_{ij}^* A_i x_{ij}}{y_{ij}^* x_{ij}}, \quad \beta_{ij} = \frac{y_{ij+1}^* x_{ij+1}}{y_{ij}^* x_{ij}} \neq 0.$$

The proof is complete.⁷⁾

2.4. We now turn to the problems raised in §1. The following theorem assures the possibility of the algorithm in Case 1.

THEOREM 3. *Let us apply the Lanczos algorithm to A with initial vectors x_1 and y_1 , and assume that Case 1 occurs after several modifications due to Cases 1-3. Namely let*

$$x_1, \dots, x_{p_1}, z_{p_1+1}, x_{p_1+2}, \dots, x_{p_2}, z_{p_2+1}, \dots, x_{p_r}, x_{p_r+1} = 0,$$

$$y_1, \dots, y_{q_1}, w_{q_1+1}, y_{q_1+2}, \dots, y_{q_2}, w_{q_2+1}, \dots, y_{q_s}, y_{q_s+1} = 0,$$

$$x_i \neq 0 \quad (1 \leq i \leq p_r, i \neq p_1+1, \dots, p_{r-1}+1),$$

$$y_j \neq 0 \quad (1 \leq j \leq q_s, j \neq q_1+1, \dots, q_{s-1}+1),$$

$$x_{p_i+1} = y_{q_j+1} = 0 \quad (1 \leq i \leq r, 1 \leq j \leq s),$$

7) The similar proof for this theorem is found in [6], but there it is not clear whether there are vectors z_{p+1}, w_{p+1} ($w_{p+1}^* z_{p+1} \neq 0$) such that they have a common grade and $y_j^* z_{p+1} = x_j^* w_{p+1} = 0$ ($1 \leq j \leq p$), in case where $x_{p+1} = y_{p+1} = 0$.

and

$$p_r = q_s = p(\text{say}).$$

Then there exists a pair of vectors z_{p+1}, w_{p+1} with a common grade such that

- (i) $z_{p+1} \in [y_1, \dots, y_{q_1}, w_{q_1+1}, y_{q_1+2}, \dots, y_{q_2}, w_{q_2+1}, \dots, y_p]^\perp,$
- (ii) $w_{p+1} \in [x_1, \dots, x_{p_1}, z_{p_1+1}, x_{p_1+2}, \dots, x_{p_2}, z_{p_2+1}, \dots, x_p]^\perp,$
- (iii) $w_{p+1}^* z_{p+1} \neq 0,$

and

(iv) the algorithm starting again from z_{p+1} and w_{p+1} can be well continued so that Case 1 occurs, i.e., so that, for some integer p_{r+1} , we have

$$x_{p_{r+1}+1} = y_{p_{r+1}+1} = 0, \quad \text{and} \quad y_i^* x_i \neq 0, \quad (p+2 \leq i \leq p_{r+1}).$$

Namely, under the above situation, the algorithm can be well continued to completion using only modifications due to Case 1.

PROOF. Let

$$\mathcal{U} = [x_1, \dots, x_{p_1}, z_{p_1+1}, x_{p_1+2}, \dots, x_{p_2}, z_{p_2+1}, \dots, x_p],$$

$$\mathcal{Q} = [y_1, \dots, y_{q_1}, w_{q_1+1}, y_{q_1+2}, \dots, y_{q_2}, w_{q_2+1}, \dots, y_p],$$

and u_1, \dots, u_{n-p} (v_1, \dots, v_{n-p}) be a basis of \mathcal{U}^\perp (\mathcal{Q}^\perp). Then the subspace \mathcal{Q}^\perp is invariant under A . In fact we have

$$y_i^*(Av) = (A^* y_i)^* v = 0 \quad (1 \leq i \leq p)$$

and

$$w_{q_j+1}^*(Av) = (A^* w_{q_j+1})^* v = 0 \quad (1 \leq j \leq s-1)$$

for any vector $v \in \mathcal{Q}^\perp$ since

$$A^* y_i, A^* w_{q_j+1} \in \mathcal{Q} \quad (1 \leq i \leq p, 1 \leq j \leq s-1).$$

Thus we may write

$$Av_i = \sum_{j=1}^{n-p} b_{ji} v_j \quad (1 \leq i \leq n-p)$$

for some scalar b_{ji} . Let B be a matrix of order $n-p$ constructed from the coefficients b_{ji} ;

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n-p} \\ b_{21} & b_{22} & \dots & b_{2n-p} \\ \dots & \dots & \dots & \dots \\ b_{n-p1} & b_{n-p2} & \dots & b_{n-p \ n-p} \end{pmatrix}.$$

Then, by virtue of Theorem 1 or 2, we can find a pair of $n-p$ dimensional vectors \tilde{z}_{p+1} and \tilde{w}_{p+1} such that B is well transformed into a block tri-diagonal matrix $L_1 \oplus \cdots \oplus L_s$ by the Lanczos algorithm starting from the pair. Let the order of L_1 be n_1 and put $p_{r+1} = p + n_1$. If we denote the iterated vectors for B by \tilde{x}_i, \tilde{y}_i , we have

$$\tilde{w}_{p+1}^* \tilde{z}_{p+1} \neq 0, \quad \tilde{y}_{p+j}^* \tilde{x}_{p+j} \neq 0 \quad (2 \leq j \leq n_1), \quad \tilde{x}_{p_{r+1}+1} = \tilde{y}_{p_{r+1}+1} = 0,$$

$$\tilde{y}_i \in [\tilde{z}_{p+1}, \tilde{x}_{p+2}, \dots, \tilde{x}_{i-1}]^\perp,$$

and

$$\tilde{x}_i \in [\tilde{w}_{p+1}, \tilde{y}_{p+2}, \dots, \tilde{y}_{i-1}]^\perp \quad (p+2 \leq i \leq p_{r+1}).$$

Let

$$X = (x_1, \dots, x_{p_1}, z_{p_1+1}, x_{p_1+2}, \dots, x_p, v_1, \dots, v_{n-p}),$$

$$Y = (y_1, \dots, y_{q_1}, w_{q_1+1}, y_{q_1+2}, \dots, y_p, u_1, \dots, u_{n-p}),$$

and

$$C = \begin{pmatrix} u_1^* v_1 & u_1^* v_2 & \cdots & u_1^* v_{n-p} \\ u_2^* v_1 & u_2^* v_2 & \cdots & u_2^* v_{n-p} \\ \dots & \dots & \dots & \dots \\ u_{n-p}^* v_1 & u_{n-p}^* v_2 & \cdots & u_{n-p}^* v_{n-p} \end{pmatrix}.$$

Then the matrix C is non-singular since X and Y are non-singular and

$$Y^* X = \left(\begin{array}{ccc|c} y_1^* x_1 & & & \\ & \ddots & & \\ & & y_p^* x_p & \\ \hline & & & C \end{array} \right).$$

Now we shall show that a pair of vectors $z_{p+1} = (v_1, \dots, v_{n-p}) \tilde{z}_{p+1}$ and $w_{p+1} = (u_1, \dots, u_{n-p}) C^{*-1} \tilde{w}_{p+1}$ is what we seek. It is clear that the conditions (i) and (ii) are satisfied since z_{p+1} and w_{p+1} are linear combinations of v_1, \dots, v_{n-p} and u_1, \dots, u_{n-p} respectively. Further we have

$$w_{p+1}^* z_{p+1} = \tilde{w}_{p+1}^* \tilde{z}_{p+1} \neq 0.$$

Next, to prove the condition (iv), we denote by x_i, y_i ($i \geq p+2$) the iterated vectors which are obtained by applying the algorithm to A with the modified vectors z_{p+1}, w_{p+1} . Then they satisfy the relations

$$x_i = (v_1, \dots, v_{n-p}) \tilde{x}_i, \quad y_i = (u_1, \dots, u_{n-p}) C^{*-1} \tilde{y}_i \quad (p+2 \leq i \leq p_{r+1}),$$

and

$$x_{p_{r+1}+1} = y_{p_{r+1}+1} = 0,$$

as is easily verified using induction on $i \geq p+2$. Let

$$Z = (x_1, \dots, x_{p_1}, z_{p_1+1}, x_{p_1+2}, \dots, x_p, z_{p+1}, x_{p+2}, \dots, x_{p_{r+1}}).$$

Then, by noting that $A(v_1, \dots, v_{n-p}) = (v_1, \dots, v_{n-p})B$, we obtain

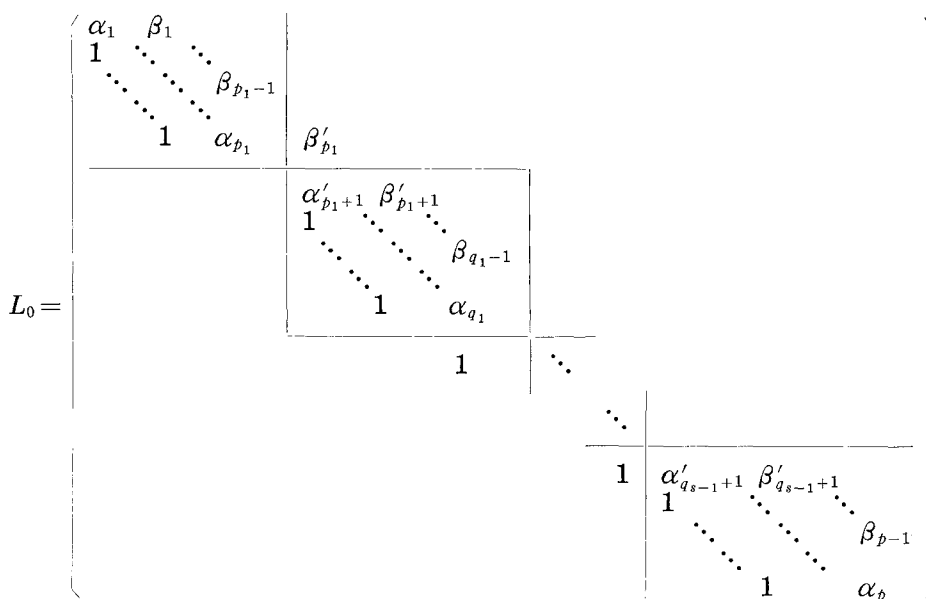
$$AZ = Z(L_0 \oplus L_1)$$

where L_0 is a (block) tri-diagonal matrix of order p such that

$$A(x_1, \dots, x_{p_1}, z_{p_1+1}, x_{p_1+2}, \dots, x_p) = (x_1, \dots, x_{p_1}, z_{p_1+1}, x_{p_1+2}, \dots, x_p)L_0,$$

and its concrete form will be shown below. This completes the proof.

.....
 Typical diagram of L_0 in case of $p_1 < q_1 < \dots < q_{s-1} < p_r (= q_s = p)$



2.5. We shall now consider the possibility of the algorithm after the modification in Case 2 or 3. First we show the following:

LEMMA 4. Let A_i ($1 \leq i \leq s$) be s Jordan block matrices of order n_i ($\sum_{i=1}^s n_i = n$) such that

$$A_i = \begin{pmatrix} \lambda_i & & & \\ & 1 & & \\ & & \lambda_i & \\ & & & \ddots \\ & & & & 1 & \\ & & & & & \lambda_i \end{pmatrix}, \quad \lambda_i \neq \lambda_j \quad (i \neq j),$$

and let $A = A_1 \oplus \dots \oplus A_s$. Corresponding to this matrix A , let $v = v_1 \oplus \dots \oplus v_s$ be an n -dimensional vector such that each subvector v_i is n_i -dimensional. Then v has the grade n with respect to $A(A^*)$ if and only if the first (last) component of each v_i is different from zero.

PROOF. Let $v_i = {}^t(v_{i1}, v_{i2}, \dots, v_{in_i})$. If, for instance, we assume that $v_{11} = 0$, then the first row of a matrix $(v, Av, \dots, A^{n-1}v)$ consists of only zero elements. Hence a set of n vectors $v, Av, \dots, A^{n-1}v$ is linearly dependent, and v can not be of grade n with respect to A ; namely, if v has the grade n with respect to A , we must have $v_{i1} \neq 0$ for every i ($1 \leq i \leq s$). In this case a simple computation on determinant shows that

$$0 \neq \det(v, Av, \dots, A^{n-1}v) = \left(\prod_{i=1}^s v_{i1} \right) \cdot \begin{vmatrix} a_1(0) & a_1(1) & \dots & a_1(n-1) \\ a_2(0) & a_2(1) & \dots & a_2(n-1) \\ \dots & \dots & \dots & \dots \\ a_s(0) & a_s(1) & \dots & a_s(n-1) \end{vmatrix}$$

where $a_i(j)$ stand for n_i -dimensional column vectors whose components consist of the first column of A_i^j . (A_i^0 is an identity matrix of order n_i .) Hence the value of the determinant on the right, denoted by Δ , is non-vanishing⁸⁾. From this it follows that $\det(v, Av, \dots, A^{n-1}v) \neq 0$ if $v_{i1} \neq 0$ ($1 \leq i \leq s$), since Δ is independent of the components of v . This proves the assertion.

LEMMA 5. Let us assume that Case 2 occur at the $p+1$ -th step (it may occur at the $p' (< p)$ th step), and modify the algorithm choosing a new vector $z_{p+1} \in [y_1, \dots, y_p]^\perp$. If $f_i(z_{p+1}, y_{p+1}, A) \neq 0$ ($1 \leq i \leq k$), then a sequence of the iterated vectors x_i, y_i ($p+2 \leq i \leq p+k$) is well defined by this modification, and we have

$$f_k(z_{p+1}, y_{p+1}, A) = (y_{p+1}^* z_{p+1}) \cdot \prod_{i=p+2}^{p+k} (y_i^* x_i)$$

where f_k is defined as in Lemma 2.

PROOF. Induction on k . Since the lemma is trivial for $k=1$, we suppose that it holds for $k-1$. Then x_i, y_i ($p+2 \leq i \leq p+k-1$) are well defined and

8) If $n_1 \geq n_2 \geq \dots \geq n_s$, we can show that

$$\Delta = \prod_{i>j} (\lambda_i - \lambda_j)^{n_i n_j}$$

by noting that

$$\det(a_s(n - n_s), a_s(n - n_s + 1), \dots, a_s(n - 1)) = \lambda_s^{n_s(n - n_s)}$$

and

$$\frac{\partial^\nu \Delta}{\partial \lambda_i^\nu} \Big|_{\lambda_i = \lambda_j} = 0 \quad (1 \leq \nu \leq n_i n_j - 1, i > j),$$

considering as a function of λ_i . Hence Lemma 4 follows from this fact. But such calculations are not necessary for our purpose.

$$f_{k-1}(z_{p+1}, y_{p+1}, A) = (y_{p+1}^* z_{p+1}) \cdot \prod_{i=2}^{k-1} (y_{p+i}^* x_{p+i}) \neq 0.$$

Hence x_{p+k}, y_{p+k} can be constructed and we have

$$\begin{aligned} x_{p+j} &= \varphi_j(A)z_{p+1} - \psi_j(A)x_p, \\ y_{p+j} &= \varphi_j(A)^* y_{p+1} \quad (2 \leq j \leq k), \end{aligned}$$

where $\varphi_j(\lambda)$ is a monic polynomial of degree $j-1$ and $\psi_j(\lambda)$ is a polynomial of degree $j-2$. Since $x_{p+1}=0$ by assumption, we have $A^q x_p \in [x_1, \dots, x_p]$ for any q and $y_{p+j}^* \psi_i(A)x_p = 0$ for any i and j ($2 \leq j \leq k$). This implies that

$$y_{p+j}^* x_{p+i} = y_{p+j}^* \varphi_i(A)z_{p+1}.$$

Moreover, $y_{p+j}^* \varphi_i(A)z_{p+1}$ is a linear combination of $y_{p+j}^* z_{p+1}, y_{p+j}^* Az_{p+1}, \dots, y_{p+j}^* A^{i-1}z_{p+1}$ with coefficient one over the last term. Thus, by elementary calculation on determinant, we have

$$\begin{aligned} (y_{p+1}^* z_{p+1}) \cdot \prod_{j=2}^k (y_{p+j}^* x_{p+j}) &= \begin{vmatrix} y_{p+1}^* z_{p+1} & \cdots & y_{p+1}^* x_{p+k} \\ y_{p+2}^* z_{p+1} & \cdots & y_{p+2}^* x_{p+k} \\ \dots & \dots & \dots \\ y_{p+k}^* z_{p+1} & \cdots & y_{p+k}^* x_{p+k} \end{vmatrix} \\ &= \begin{vmatrix} y_{p+1}^* z_{p+1} & y_{p+1}^* Az_{p+1} & \cdots & y_{p+1}^* A^{k-1}z_{p+1} \\ y_{p+1}^* Az_{p+1} & y_{p+1}^* A^2z_{p+1} & \cdots & y_{p+1}^* A^kz_{p+1} \\ \dots & \dots & \dots & \dots \\ y_{p+1}^* A^{k-1}z_{p+1} & y_{p+1}^* A^kz_{p+1} & \cdots & y_{p+1}^* A^{2k-2}z_{p+1} \end{vmatrix} \\ &= f_k(z_{p+1}, y_{p+1}, A). \quad \text{Q.E.D.} \end{aligned}$$

LEMMA 6. If $x(y)$ is a given vector of grade p with respect to A (A^*), then we have

$$f_k(x, y, A) \neq 0 \quad (1 \leq k \leq p)$$

considering as a function of the components of a vector $y(x)$.

PROOF. Obviously, we may assume that $A = \sum_{i=1}^s \oplus A_i$ with

$$A_i = \left(\begin{array}{c|c|c} \left. \begin{array}{ccc} \lambda_i & & \\ 1 & \ddots & \\ & & 1 & \lambda_i \end{array} \right\} n_1(i) & & \\ \hline & \left. \begin{array}{ccc} \lambda_i & & \\ 1 & \ddots & \\ & & 1 & \lambda_i \end{array} \right\} n_2(i) & & \\ \hline & & & \ddots \end{array} \right), \quad \lambda_i \neq \lambda_j \quad (i \neq j).$$

Then we put

$$x = x(1) \oplus \cdots \oplus x(s), \quad y = y(1) \oplus \cdots \oplus y(s),$$

where $x(i)$, $y(i)$ correspond to A_i , and

$$\begin{aligned} x(i) &= {}^t(\xi_{11}(i), \xi_{12}(i), \dots, \xi_{1n_1(i)}(i), \xi_{21}(i), \dots, \xi_{2n_2(i)}(i), \dots), \\ y(i) &= {}^t(\eta_{11}(i), \eta_{12}(i), \dots, \eta_{1n_1(i)}(i), \eta_{21}(i), \dots, \eta_{2n_2(i)}(i), \dots). \end{aligned}$$

In order to prove the lemma, it is sufficient to show that, for any k with $1 \leq k \leq p$, we can construct a vector y such that $f_k(x, y, A) \neq 0$. Now, given i , define an integer $h_j(i)$ for each j , as follows:

$$h_j(i) = \begin{cases} q & \text{if } \xi_{jq+1}(i) \neq 0 \text{ and } \xi_{jt}(i) = 0 \text{ for } t \leq q, \\ 0 & \text{if } \xi_{j1}(i) \neq 0. \end{cases}$$

Next, let

$$d_i = \max_j \{n_j(i) - h_j(i)\}$$

and

$$N_i = A_i - \lambda_i I_i$$

where I_i is an identity matrix of the same order as of A_i . Then, for any k such that $1 \leq k \leq p$, we have

$$\begin{aligned} k \leq p &= \text{rank}(x, Ax, \dots, A^{p-1}x) \\ &\leq \sum_{i=1}^s \text{rank}(x(i), A_i x(i), \dots, A_i^{p-1}x(i)) \\ &= \sum_{i=1}^s \text{rank}(x(i), N_i x(i), \dots, N_i^{p-1}x(i)) \\ &= \sum_{i=1}^s d_i. \end{aligned}$$

Hence it is possible to select non-negative integer k_i so that $k_i \leq d_i$ and $\sum_{i=1}^s k_i = k$. Without loss of generality we may assume that

$$d_i = n_1(i) - h_1(i) \quad (1 \leq i \leq s).$$

Then, putting $h_i = h_1(i)$ in order to simplify the notation, we define k_i -dimensional vectors $\tilde{x}(i)$, $\tilde{y}(i)$ and k_i -square matrices \tilde{A}_i as follows:

$$\begin{aligned} \tilde{x}(i) &= {}^t(\xi_{1h_i+1}(i), \xi_{1h_i+2}(i), \dots, \xi_{1h_i+k_i}(i)), \\ \tilde{y}(i) &= {}^t(\eta_{1h_i+1}(i), \eta_{1h_i+2}(i), \dots, \eta_{1h_i+k_i}(i)), \\ \tilde{A}_i &= A_i(\xi_{h_i+1, h_i+2}, \dots, \xi_{h_i+k_i, h_i+k_i}). \end{aligned}$$

Now, for the vector x given above, consider a vector y such that

$$\eta_{jt}(i) = 0 \quad (j \neq 1, \text{ or } j=1 \text{ and } t > h_i + k_i)$$

and

$$\eta_{1h_i+k_i}(i) \neq 0.$$

Then this vector y satisfies $f_k(x, y, A) \neq 0$. In fact, if we put

$$\tilde{x} = \tilde{x}(1) \oplus \dots \oplus \tilde{x}(s), \quad \tilde{y} = \tilde{y}(1) \oplus \dots \oplus \tilde{y}(s),$$

and

$$\tilde{A} = \tilde{A}_1 \oplus \dots \oplus \tilde{A}_s,^{9)}$$

then, by Lemma 4, a set of k -dimensional vectors $\tilde{x}, \tilde{A}\tilde{x}, \dots, \tilde{A}^{k-1}\tilde{x}(\tilde{y}, \tilde{A}^*\tilde{y}, \dots, \tilde{A}^{*k-1}\tilde{y})$ is linearly independent since the first (last) component $\xi_{1h_i+k_i}(i)$ ($\eta_{1h_i+k_i}(i)$) of the vector $\tilde{x}(i)$ ($\tilde{y}(i)$) is different from zero for each i . Also it is clear that $y^*A^i x = \tilde{y}^*\tilde{A}^i \tilde{x}$ for any non-negative integer i . Thus we obtain

$$\begin{aligned} f_k(x, y, A) &= f_k(\tilde{x}, \tilde{y}, \tilde{A}) \\ &= \begin{vmatrix} \tilde{y}^* \\ \tilde{y}^* \tilde{A} \\ \dots \\ \tilde{y}^* \tilde{A}^{k-1} \end{vmatrix} \cdot |\tilde{x}, \tilde{A}\tilde{x}, \dots, \tilde{A}^{k-1}\tilde{x}| \neq 0, \end{aligned}$$

i.e., $f_k(x, y, A) \neq 0 \quad (1 \leq k \leq p)$,

which establishes the assertion. Q.E.D.

We are now in a position to prove the following:

THEOREM 4. *Let us apply the Lanczos algorithm to an n -square matrix A with initial vectors x_1 and y_1 . And assume that Case 2 occurs after several modifications due to Cases 1-3. Namely, let the iterated vectors be obtained as follows:*

$$\begin{aligned} (4) \quad & x_1, \dots, x_{p_1}, z_{p_1+1}, x_{p_1+2}, \dots, x_{p_2}, z_{p_2+1}, \dots, x_{p_r}, x_{p_r+1} = 0, \\ & y_1, \dots, y_{q_1}, w_{q_1+1}, y_{q_1+2}, \dots, y_{q_2}, w_{q_2+1}, \dots, y_{q_s}, y_{q_s+1} \neq 0, \\ & x_{p_i+1} = 0 \quad (1 \leq i \leq r), \quad y_{q_j+1} = 0 \quad (1 \leq j \leq s-1), \end{aligned}$$

where $p_r = q_s$. Then there exists a vector z_{p_r+1} such that

- (i) $z_{p_r+1} \in [y_1, \dots, y_{q_1}, w_{q_1+1}, y_{q_1+2}, \dots, w_{q_{s-1}+1}, \dots, y_{p_r}]^\perp$,
- (ii) $y_{p_r+1}^* z_{p_r+1} \neq 0$,

9) If $d_i = 0$ for some i , then $k_i = 0$ and $\tilde{x}(i)$, $\tilde{y}(i)$ and \tilde{A}_i do not appear.

and

(iii) the algorithm starting again from z_{p_r+1} and y_{p_r+1} can be well continued to completion.

In particular, if A and the vectors in (4) are all real, the vector z_{p_r+1} can be taken as a real vector.

PROOF. For convenience sake, let $p_r = q_s = p$ and let \mathcal{U}, \mathcal{Q} be defined as in the proof of Theorem 3. Then it is clear that there exists a vector z_{p+1} such that $z_{p+1} \in \mathcal{Q}^\perp$ and $y_{p+1}^* z_{p+1} \neq 0$, since the union of a set of the vectors $x_1, \dots, x_{p_1}, z_{p_1+1}, \dots, x_p$ and a basis of \mathcal{Q}^\perp spans the whole space. If either one of Cases 1-3 occurs after starting again from a pair of vectors z_{p+1} and y_{p+1} , we can continue the process by the modification as is explained in §1. Therefore, in order to prove the theorem, it is sufficient to show that a vector z_{p+1} satisfying (i) and (ii) can be chosen so that Case 4 does not occur. On the contrary, suppose that Case 4 occurs for any choice of a vector z_{p+1} satisfying the condition (i) and (ii). Then there exists a positive integer $k = k(z_{p+1}) (\geq 2)$ depending on z_{p+1} such that

$$(5) \quad \begin{aligned} & y_{p+j}^* x_{p+j} \neq 0 \quad (2 \leq j \leq k-1), \\ & y_{p+k}^* x_{p+k} = 0, \quad x_{p+k} \neq 0 \quad \text{and} \quad y_{p+k} \neq 0, \end{aligned}$$

where $x_{p+j}, y_{p+j} (2 \leq j \leq k)$ denote the iterated vectors obtained by the algorithm starting again from the vectors z_{p+1} and y_{p+1} . Since $k < n - p$,

$$q = \max_{\substack{z_{p+1} \in \mathcal{Q}^\perp \\ y_{p+1}^* z_{p+1} \neq 0}} k(z_{p+1})$$

exists and we can find a vector z_{p+1} such that the situation (5) happens at $k = q$. Then the grade of y_{p+1} with respect to A^* is not less than q since

$$q = \text{rank}(y_{p+1}, \dots, y_{p+q}) = \text{rank}(y_{p+1}, A^* y_{p+1}, \dots, A^{*q-1} y_{p+1}).$$

Hence, by Lemma 6, we have

$$f_i(x, y_{p+1}, A) \neq 0 \quad (1 \leq i \leq q)$$

considering as a function of the components of x . This implies the existence of a vector z such that

$$f_i(z, y_{p+1}, A) \neq 0 \quad (1 \leq i \leq q).$$

Since the whole space is the direct sum of the space \mathcal{U} and \mathcal{Q}^\perp , the vector z can be written uniquely in the form

$$z = u + v \quad (u \in \mathcal{U}, v \in \mathcal{Q}^\perp).$$

We put $z_{p+1} = v$. Since, as is easily seen, the space \mathcal{U} is invariant under A , we have for any positive integer h

$$A^h u \in \mathcal{U}, \quad y_{p+1}^* A^h u = 0,$$

and

$$y_{p+1}^* A^h z_{p+1} = y_{p+1}^* A^h z.$$

Therefore we have

$$f_i(z_{p+1}, y_{p+1}, A) = f_i(z, y_{p+1}, A) \neq 0 \quad (1 \leq i \leq q).$$

On the other hand, since the vector z_{p+1} satisfies the conditions (i) and (ii), we can continue the process starting again from a pair of the vectors z_{p+1} and y_{p+1} . Then, by Lemma 5, the following holds:

$$f_q(z_{p+1}, y_{p+1}, A) = (y_{p+1}^* z_{p+1}) \cdot \prod_{j=2}^q (y_{p+j}^* x_{p+j})$$

where x_{p+j}, y_{p+j} represent the iterated vectors constructed by this algorithm. Hence we must have $y_{p+q}^* x_{p+q} \neq 0$, which contradicts to the maximality of q . Thus there exists a vector z_{p+1} such that $z_{p+1} \in \mathbb{C}^{\mathcal{D}^+}$, $y_{p+1}^* z_{p+1} \neq 0$ and Case 4 does not occur. Especially, if A and the iterated vectors in (4) are all real, it is clear that the vector z_{p+1} can be chosen as a real vector. The proof is complete.

As a special case of Theorem 4 we have

THEOREM 5. *For a given non-zero vector $x(y)$, there exists a vector $y(x)$ such that the Lanczos algorithm starting from $x_1 = x$ and $y_1 = y$ can be well continued to completion. Especially, in Theorem 2, one of the vectors x_1, y_1 can be chosen arbitrarily as long as it has the grade m .*

2.6. As is well known, if A is hermitian (or real symmetric), the algorithm is well continued to completion, starting from any common initial vector $x_1 = y_1 = x$. This is not true in general, even for normal matrices as the following simple example shows:

Consider a normal matrix

$$A = \begin{pmatrix} i & i & 0 \\ i & i & 0 \\ 0 & 0 & \sqrt{3} + i \end{pmatrix}.$$

If we choose a vector $x = {}^t(\sqrt{2}/3, 0, 1/\sqrt{3})$ as a common initial vector, then we have

$$x_2 = Ax - \frac{x^* Ax}{x^* x} x = {}^t(-\sqrt{2}/3, i\sqrt{2}/3, 2/3),$$

$$y_2 = A^*x - \frac{x^*A^*x}{x^*x}x = {}^t(-\sqrt{2}/3, -i\sqrt{2}/3, 2/3),$$

and

$$y_2^*x_2 = 0.$$

This example raises a question whether, for a given matrix A , there always exists a vector x such that the algorithm starting from $x_1 = y_1 = x$ can be well continued to completion. Fortunately the answer is affirmative. Namely, as another refinement of Theorem 2 (Rutishauser), we obtain

THEOREM 6. *In Theorem 2, x_1 and y_1 can be taken as the same vector. Namely we can find a vector x of grade m with respect to both A and A^* so that the algorithm starting from $x_1 = y_1 = x$ may be well continued to completion using only modifications due to Case 1. If A is real, the vector x may be taken as a real vector and the algorithm is possible in the realm of reals.*

PROOF. Let T and A_1 be the matrices defined in the proof of Lemma 2. We denote by J_i ($1 \leq i \leq s$) the i -th Jordan block matrices of order n_i appeared in A_1 ; i.e., $A_1 = \sum_{i=1}^s \oplus J_i$. Then it will be shown that

$$(6) \quad f_k(x, T^*Tx, T^{-1}AT) \neq 0 \quad (1 \leq k \leq m)$$

considering as a function of the components of a vector x . To prove this, we take a positive integer k with $1 \leq k \leq m$ and r ($\leq s$) positive integers k_i such that $k_i \leq n_i$ and $\sum_{i=1}^r k_i = k$. Further we put

$$x = \left(\sum_{i=1}^s \oplus x(i) \right) \oplus {}^t \overbrace{(0, \dots, 0)}^{n-m}, \quad T^*Tx = \left(\sum_{i=1}^s \oplus y(i) \right) \oplus {}^t \overbrace{(*, \dots, *)}^{n-m},$$

where $x(i)$, $y(i)$ are n_i -dimensional vectors and

$$x(i) \begin{cases} = {}^t(0, \dots, 0, \xi_{i1}, \dots, \xi_{ik_i}) & (1 \leq i \leq r) \\ = {}^t(0, \dots, 0) & (r+1 \leq i \leq s), \end{cases}$$

$$y(i) = {}^t(*, \dots, *, \eta_{i1}, \dots, \eta_{ik_i}) \quad (1 \leq i \leq r).$$

Since T^*T is positive definite, each component η_{ij} is a non-trivial function of $\xi_{11}, \dots, \xi_{1k_1}, \dots, \xi_{r1}, \dots, \xi_{rk_r}$. Hence we can find k numbers ξ_{hl} such that $\eta_{ij} \neq 0$ ($1 \leq i \leq r, 1 \leq j \leq k_i$). By an ordinary argument of continuity, we may assume that $\xi_{hl} \neq 0$ ($1 \leq h \leq r, 1 \leq l \leq k_i$). Then, in the same way as in the proof of Lemma 6, we obtain $f_k(x, T^*Tx, T^{-1}AT) = f_k(\hat{x}, \hat{y}, \hat{A}) \neq 0$, where

$$\hat{x} = {}^t(\xi_{11}, \dots, \xi_{1k_1}, \dots, \xi_{r1}, \dots, \xi_{rk_r}),$$

$$\hat{y} = {}^t(\eta_{11}, \dots, \eta_{1k_1}, \dots, \eta_{r1}, \dots, \eta_{rk_r}),$$

and

$$\hat{A} = \sum_{i=1}^r \oplus J_i(n_i - k_i + 1, n_i - k_i + 2, \dots, n_i).$$

Since k is an arbitrary integer such that $1 \leq k \leq m$, this establishes (6). Therefore we can find a vector \tilde{x} such that $f_k(\tilde{x}, T^*T\tilde{x}, T^{-1}AT) \neq 0$ ($1 \leq k \leq m$). Then the grade of \tilde{x} with respect to $T^{-1}AT$ is clearly m and the algorithm for $T^{-1}AT$ starting from initial vectors $\tilde{x}_1 = \tilde{x}$ and $\tilde{y}_1 = T^*T\tilde{x}$ can be continued to the m -th step:

$$\tilde{y}_i^* \tilde{x}_i \neq 0 \quad (1 \leq i \leq m), \quad \tilde{x}_{m+1} = \tilde{y}_{m+1} = 0$$

where \tilde{x}_i and \tilde{y}_i denote the i -th iterated vectors applied to $T^{-1}AT$. Therefore, by Theorem 3, the algorithm can be well continued to completion. By Lemma 1, this implies that the algorithm for A starting from common initial vectors $x_1 = y_1 = T\tilde{x}_1$ can be well continued to completion using only modifications for Case 1. Evidently the vector $T\tilde{x}_1$ has the grade m with respect to both A and A^* . The remaining part is clear. Q.E.D.

2.7. Computational procedure. So far, we discussed the possibility of the Lanczos algorithm from theoretical point of view. Now, according to the results obtained there, a computational procedure of the algorithm can be formulated as follows:

Step 1. Let x_1 and y_1 be a pair of vectors which is chosen arbitrarily or according to any criterion, and start the algorithm.

Step 2. If Case 4 first occurs on the way, we choose a new vector x'_1 and begin again with a pair of vectors x'_1 and y_1 .

Step 3. If either one of Cases 1-3 occurs on the way, we modify the procedure according to the rule stated in §1, and continue the iteration.

Step 4. Proceeding in this way, if Case 4 occurs after several modification due to Cases 1-3, we go back to the latest modification and begin again from there replacing the modified vector by a new one. Namely, if the latest modification is due to Case 1 at the $p+1$ -th step, we may replace only one of the vectors z_{p+1} and w_{p+1} by a new vector; similarly, if it is due to Case 2 (3) at the $p+1$ -th step, it is sufficient to replace the vector $z_{p+1}(w_{p+1})$ by a new vector $z'_{p+1}(w'_{p+1})$.

In the above procedure, if A is a real matrix, the algorithm can be done in the realm of real, i.e., vectors $x_1, y_1, z_{p+1}, w_{p+1}$, etc. may be taken as real vectors. At any rate, theoretically, the algorithm is always possible by the above procedures as Theorems 3,4 and 5 guarantee. Further, by Theorem 6, we may replace "Step 1" by the following:

Step 1'. Choosing any non-zero vector x , start the algorithm from $x_1 = y_1 = x$.

2.8. Geometric interpretation. Let A be a non-derogatory matrix of order n , and S be an n -dimensional (complex or real according as A is complex or real) affine space. For each positive integer k , let V_k be an algebraic variety defined by the equation $f_k(x, y, A)=0$. Considering an n -dimensional vector as a point of the space S , we shall call a pair of the initial vectors leading to one of Cases 1-4 as a breakdown point in a space $S \times S$. Then a set of all the breakdown points forms an algebraic variety $V = \bigcup_{k=1}^n V_k$ in $S \times S$ defined by $\prod_{k=1}^n f_k(x, y, A)=0$ since

$$f_k(x, y, A) = \prod_{i=1}^k (y_i^* x_i)$$

or

$$f_k(z_{p+1}, y_{p+1}, A) = (z_{p+1}^* y_{p+1}) \cdot \prod_{j=2}^k (y_{p+j}^* x_{p+j}),$$

etc. by Lemma 5. Thus the results (Theorems 2-6 and Lemma 6) suggest the following geometric interpretation for the possibility of the Lanczos algorithm.

THEOREM. *Let A, S , and V_i be defined as above.*

(i) *A set of all the breakdown points forms an algebraic variety $V = \bigcup_{i=1}^n V_i$ in $S \times S$. And there exists a point P of $S \times S$ such that $P \notin V$.*

(ii) *For any point $x (\neq 0) \in S$ having the grade p with respect to A , we have $x \times S \not\subseteq \bigcup_{i=1}^p V_i$ and certainly $x \times S \subseteq V_{p+1}$. Analogously we have $S \times y \not\subseteq \bigcup_{i=1}^p V_i$ and $S \times y \subseteq V_{p+1}$ for any point $y (\neq 0) \in S$ having the grade p with respect to A^* .*

(iii) *The diagonal in $S \times S$ is not contained in V .*

Appendix. The eigenvalues of tri-diagonal matrices

In this appendix, we investigate some properties concerning the eigenvalues of tri-diagonal matrices, in connection with the Lanczos algorithm. Let $A=(a_{ij})$ be an upper Hessenberg matrix of order n . If $a_{i+1i}=0$ for some i , the eigenvalue problem for A is reduced to that of lower order. Hence there is no loss of generality even if we assume that $a_{i+1i} \neq 0$ for any i . Then the following lemma is clear from the theory on elementary divisors since the elementary divisors e_i satisfy $e_i=1$ ($1 \leq i \leq n-1$). But we give here another elementary proof for the sake of completeness.

LEMMA. *Let $A=(a_{ij})$ be an upper Hessenberg matrix with $a_{i+1i} \neq 0$*

($1 \leq i \leq n-1$). Then the eigenvalues of A are distinct if and only if A is diagonalizable.

PROOF. let $\lambda_1, \lambda_2, \dots, \lambda_k$ be k numbers and consider a matrix

$$\tilde{A} = (A - \lambda_1 I)(A - \lambda_2 I) \dots (A - \lambda_k I),$$

where I denotes the identity matrix of order n . If $k < n$, then $\tilde{A} \neq 0$ since, as is easily seen, the $(k+1, 1)$ element of \tilde{A} is $\prod_{i=1}^k a_{i+1i} \neq 0$. Therefore the degree of the minimal polynomial for A must be n . Hence the eigenvalues of A are distinct if A is diagonalizable. The converse is clear. Q.E.D.

Since a tri-diagonal matrix is a special case of the Hessenberg matrix, we obtain from the lemma

THEOREM A.1. Let

$$(1) \quad A = \begin{pmatrix} b_1 & c_1 & & & \\ a_1 & b_2 & c_2 & & \\ & \ddots & \ddots & \ddots & \\ & a_{n-2} & b_{n-1} & c_{n-1} & \\ & & a_{n-1} & b_n & \end{pmatrix}$$

where a_i and c_i are real and $a_i c_i > 0$ ($1 \leq i \leq n-1$). Then we have the following:

(i) The imaginary part of any eigenvalue λ of A satisfies

$$\min_{1 \leq i \leq n} \text{Im}(b_i) \leq \text{Im}(\lambda) \leq \max_{1 \leq i \leq n} \text{Im}(b_i).$$

(ii) If b_i are real, the eigenvalues of A are real and simple.

(iii) If b_i are all real, exactly one eigenvalue of A ($1 \leq i \leq n-1$) lies between any two eigenvalues of A .

(REMARK. The properties (ii) and (iii) are well known in connection with Sturm sequence, but, as is shown below, we can give a unified treatment.)

PROOF. As is well known, by diagonal matrix D , we can transform A into

$$D^{-1}AD = \begin{pmatrix} b_1 & \sqrt{a_1 c_1} & & & \\ \sqrt{a_1 c_1} & b_2 & \sqrt{a_2 c_2} & & \\ & \ddots & \ddots & \ddots & \\ & \sqrt{a_{n-2} c_{n-2}} & b_{n-1} & \sqrt{a_{n-1} c_{n-1}} & \\ & & \sqrt{a_{n-1} c_{n-1}} & b_n & \end{pmatrix}.$$

Hence, if b_i are real, $D^{-1}AD$ is real symmetric and diagonalizable. There-

fore the eigenvalues of $D^{-1}AD$ (and A) are real and distinct by the lemma. This proves (ii). (iii) is a consequence of a direct application of the separation theorem for the real symmetric matrix $D^{-1}AD$. Now we shall show (i). Let $x = {}^t(\xi_1, \dots, \xi_n)$ be a unit eigenvector for $D^{-1}AD$ corresponding to an eigenvalue λ . Then we have

$$\lambda = x^*D^{-1}ADx = \sum_{i=1}^n b_i |\xi_i|^2 + \sum_{i=1}^{n-1} \sqrt{a_i c_i} (\xi_i \xi_{i+1} + \xi_i \bar{\xi}_{i+1}).$$

Hence we obtain

$$\text{Im}(\lambda) = \text{Im}(\sum_{i=1}^n b_i |\xi_i|^2) = \sum_{i=1}^n \text{Im}(b_i) |\xi_i|^2.$$

Thus the inequality (i) follows. Q.E.D.

As a dual for Theorem A.1, we obtain

THEOREM A.2. *In the tri-diagonal matrix (1), let a_i, c_i be real and $a_i c_i < 0$ ($1 \leq i \leq n-1$). Then we have the following:*

(i) *The real part of eigenvalue λ of A satisfies*

$$\min_{1 \leq i \leq n} \text{Re}(b_i) \leq \text{Re}(\lambda) \leq \max_{1 \leq i \leq n} \text{Re}(b_i).$$

(ii) *If $b_i = 0$ ($1 \leq i \leq n$), the eigenvalues of A are pure imaginary (admitting zero) and simple.*

PROOF. It is sufficient to consider a diagonal matrix $D = \text{diag}(1, \sqrt{-a_1/c_1}, \dots, \sqrt{\prod_{i=1}^{n-1} (-a_i/c_i)})$ and $D^{-1}AD$. Q.E.D.

COROLLARY (ARSCOTT [1]). *If the matrix A in (1) is real and $a_i c_i < 0$ ($1 \leq i \leq n-1$), then all the real eigenvalues of A lie between the least and greatest of the b_i , these values included.*

The similar results hold for a certain type of infinite tri-diagonal matrix. Let X be a separable infinite dimensional complex Hilbert space. And let A be a linear operator of X into itself. If A admits an infinite tridiagonal matrix representation

$$(2) \quad \begin{pmatrix} b_1 & c_1 & & & \\ a_1 & b_2 & c_2 & & \\ & \ddots & \ddots & \ddots & \\ & & a_{n-1} & b_n & c_n \\ & & & \ddots & \ddots \end{pmatrix}$$

with respect to some orthonormal basis of X , and $a_n, b_n, c_n \rightarrow 0$ ($n \rightarrow \infty$), then A is compact. Hence all the eigenvalues of A are approximated by the eigen-

values of finite matrix

$$A_n = \begin{pmatrix} b_1 & c_1 & & & \\ a_1 & b_2 & c_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & c_{n-1} \\ & & & a_{n-1} & b_n \end{pmatrix}$$

([4] Lemma XI. 9.5). On the other hand, let $\lambda_i(n)$ ($|\lambda_1(n)| \geq |\lambda_2(n)| \geq \dots$) be the eigenvalues of A_n , each arranged according to a certain rule. Then, for every i , any limit point of $\{\lambda_i(n)\}_{n=1}^\infty$ is a point of the spectrum of A (see [12]). Therefore Theorems A.1 and A.2 are transformed respectively as follows:

THEOREM A.1'. *Let A be an operator of X into itself and admit a matrix representation (2) with respect to some orthonormal basis of X . If a_i and c_i are real and $a_i c_i > 0$ for every i , then we have the following:*

- (i) *The imaginary part of any eigenvalue λ of A satisfies*

$$\inf_i \text{Im}(b_i) \leq \text{Im}(\lambda) \leq \sup_i \text{Im}(b_i).$$

- (ii) *If b_i are real, the eigenvalues of A are real.*

THEOREM A.2'. *Let A be an operator defined as in Theorem A.1'. If a_i and c_i are real and $a_i c_i < 0$ for every i , then we have the following:*

- (i) *The real part of any eigenvalue λ of A satisfies*

$$\inf_i \text{Re}(b_i) \leq \text{Re}(\lambda) \leq \sup_i \text{Re}(b_i).$$

- (ii) *If $b_i = 0$ for every i , the eigenvalues of A are pure imaginary (admitting zero).*

Acknowledgement. The writer wishes to express his hearty gratitude to Professor Y. Nakai for his constant guidance and encouragement during this investigation.

References

- [1] Arscott, F. M.: *Latent roots of tri-diagonal matrices*, Proc. Edinburgh Math. Soc., 12 (1961), Edinburgh Math. Notes No. 44, 5-7.
- [2] Brooker, R. A. and F. H. Sumner: *The method of Lanczos for calculating the characteristic roots and vectors of a real symmetric matrix*, Proc. Instn elect. Engrs B., 103 (1956), Supple. 114-119.
- [3] Causey, R. L. and R. T. Gregory: *On Lanczos algorithm for tri-diagonalizing matrices*, SIAM Review, 3 (1961), 322-328.
- [4] Dunford, N. and J. T. Schwartz: *Linear operators. Part II*. Interscience Publishers, 1963.
- [5] Gregory, R. T.: *Results using Lanczos' method for finding eigenvalues of arbitrary matrices*, J. Soc. Indust. Appl. Math., 6 (1958), 182-188.
- [6] Householder, A. S.: *The theory of matrices in numerical analysis*, Blaisdell Publishing Company, 1964.

- [7] Lanczos, C.: *An iteration for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Nat. Bur. Standards, **45** (1950), 255-282.
- [8] Rutishauser, H.: *Beitrage zur Kenntniss des Biorthogonalisierungs-Algorithmus von Lanczos*, Z. Angew. Math. Physik, **4** (1953), 35-56.
- [9] Wilkinson, J. H.: *The calculation of eigenvectors by the method of Lanczos*, Comput. J., **1** (1958), 148-152.
- [10] ———, *The algebraic eigenvalue problem*, Clarendon Press. Oxford, 1965.
- [11] Yamamoto, T.: *On the characteristic roots of Hessenberg matrices* (Japanese), Jyôho-Syori, **8** (1967), 154-156.
- [12] ———, *On the eigenvalues of compact operators in a Hilbert space*, Numer. Math., **11** (1968), 211-219.

*Department of Mathematics
Faculty of General Education
Hiroshima University*