ON TRACE ZERO MATRICES

by

V.S. Sunder

In this note, we shall try to present an elementary proof of a couple of closely related results which have both proved quite useful, and also indicate possible generalisations. The results we have in mind are the following facts:

(a) A complex $n \times n$ matrix A has trace 0 if and only if it is expressible in the form $A = PQ - QP$ for some P, Q.

(b) The *numerical range* of a bounded linear operator T on a complex Hilbert space $\mathcal{H}$, which is defined by

$$W(T) = \{\langle Tx, x\rangle : x \in \mathcal{H},\ ||x|| = 1\}\ ,$$

is a **convex** set in $\mathbb{C}$.[1]

We shall attempt to make the treatment easy-paced and self-contained. (In particular, all the terms in 'facts (a) and (b)' above will be described in detail.) So we shall begin with an introductory section pertaining to matrices and inner product spaces. This introductory section may be safely skipped by those readers who may be already acquainted with these topics; it is intended for those readers who have been denied the pleasure of these acquaintances.

1 Matrices and inner-product spaces

The collection $M_{m\times n}(\mathbb{C})$ of complex $m\times n$ matrices has a natural structure of a **complex vector space** in the sense that if $A = ((a_{ij})), B = ((b_{ij})) \in M_{m\times n}(\mathbb{C})$ and $\lambda \in \mathbb{C}$, we may define the *linear combination* $\lambda A + B \in M_{m\times n}(\mathbb{C})$ to be the matrix with (i, j)-th entry given by $\lambda a_{ij} + b_{ij}$. (The 'zero' of this vector space

[1] This result is known - see [H] - as the Toeplitz-Hausdorff theorem; in the statement of the theorem, we use standard set-theoretical notation, where by $x \in S$ means that x is an element of the set S.

is the $m \times n$ matrix all of whose entries are 0; this 'zero matrix' will be denoted simply by 0.)

Given two matrices whose 'sizes are suitably compatible', they may be multiplied. The product AB of two matrices A and B is defined only if there are integers m, n, p such that $A = ((a_{ik})) \in M_{m\times n}$, $B = ((b_{kj})) \in M_{n\times p}$; in that case $AB \in M_{m\times p}$ is defined as the matrix $((c_{ij}))$ given by

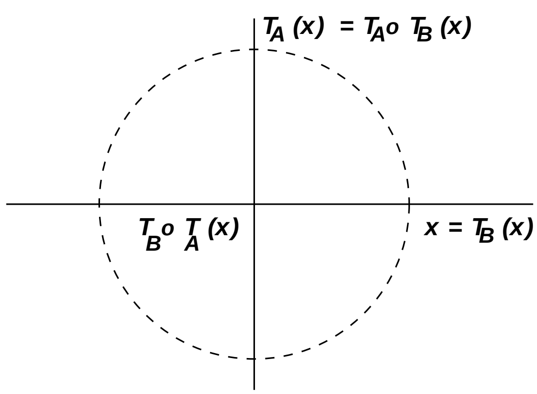
$$c_{ij} \;=\; \sum_{k=1}^{n} a_{ik} b_{kj} \;. \qquad (1.1)$$

Unlike the case of usual numbers, matrix-multiplication is not 'commutative'. For instance, if we set

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \;,\; B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \;, \qquad (1.2)$$

then it may be seen that $AB \neq BA$.

The way to think about matrices and understand matrix-multiplication is *geometrically*. When viewed properly, the reason for the validity of the example of the previous paragraph is this: if T_A denotes the operation of 'counterclockwise rotation of the plane by 90^o', and if T_B denotes 'projection onto the x-axis', then $T_A \circ T_B$, the result of doing T_B first and then T_A, is not the same as $T_B \circ T_A$, the result of doing T_A first and then T_B. (For instance, if $x = (1, 0)$, then $T_B(x) = x, T_A(x) = T_A \circ T_B(x) = (0, 1)$ while $T_B \circ T_A(x) = (0, 0)$.)



Let us see how this 'algebra-geometry' nexus goes. The correspondence

$$\mathbf{z} = (z_1, z_2, \cdots, z_n) \leftrightarrow \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} = \hat{\mathbf{z}} \tag{1.3}$$

sets up an identification between $\mathbb{C}^n$ and $M_{n\times 1}(\mathbb{C})$, which is an 'isomorphism of complex vector spaces' - in the sense that

$$\widehat{\lambda \mathbf{z} + \mathbf{z}'} = \lambda \hat{\mathbf{z}} + \hat{\mathbf{z}'}$$

Now, if $A \in M_{m\times n}(\mathbb{C})$, consider the mapping $T_A : \mathbb{C}^n \to \mathbb{C}^m$ which is defined by the requirement that if $\mathbf{z} \in \mathbb{C}^n$,then

$$\widehat{T_A(z)} = A\hat{\mathbf{z}} \tag{1.4}$$

where $A\hat{\mathbf{z}}$ denotes the matrix product of the $m \times n$ matrix A and the $n \times 1$ matrix $\hat{\mathbf{z}}$. It is then not hard to see that T_A is a *linear transformation from* $\mathbb{C}^n$ *to* $\mathbb{C}^m$: i.e., T_A satisfies the algebraic requirement[2]that

$$T_A(\lambda x + y) = \lambda T_A(x) + T_A(y) \ \text{ for all } \ x, y \in \mathbb{C}^n \ .$$

The importance of matrices stems from the fact that the converse statement is true; i.e., if T is a linear transformation from $\mathbb{C}^n$ to $\mathbb{C}^m$, then there is a unique matrix $A \in M_{m\times n}(\mathbb{C})$ such that $T = T_A$. This is an easy exercise and, we indeed have a bijective correspondence between $M_{m\times n}(\mathbb{C})$ and the collection $\mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ of linear transformations from $\mathbb{C}^n$ to $\mathbb{C}^m$. Note that the matrix corresponding to the linear transformation T is obtained by taking the j-th column as the (matrix of coefficients of the) image under T of the j-th standard basis vector. Thus, the transformation of $\mathbb{C}^2$ corresponding to 'counter-clockwise rotation by 90^o' is seen to map $e_1^{(2)}$ to $e_2^{(2)}$, and $e_2^{(2)}$ to $-e_1^{(2)}$, and the

[2]This algebraic requirement is equivalent, under mild additional conditions, to the geometric requirement that the mapping preserves 'collinearity': i.e., if x, y, z are three points in $\mathbb{C}^n$ which lie on a straight line, then the points Tx, Ty, Tz also lie on a straight line.

associated matrix is the matrix A of eqn. (1.2). (The reader is urged to check similarly that the matrix B of eqn. (1.2) does indeed correspond to 'perpendicular projection onto the x-axis'.)

Finally, if $A = ((a_{ik})) \in M_{m\times n}(\mathbb{C})$ and $B = ((b_{kj})) \in M_{n\times p}(\mathbb{C})$, then we have $T_A : \mathbb{C}^n \to \mathbb{C}^m$ and $T_B : \mathbb{C}^p \to \mathbb{C}^n$, and consequently 'composition' yields the map $T_A \circ T_B : \mathbb{C}^p \to \mathbb{C}^m$. A moment's reflection on the prescription (contained in the second sentence of the previous paragraph) for obtaining the matrix corresponding to the composite map $T_A \circ T_B$ shows the following: multiplication of matrices is defined the way it is, precisely because we have:

$$T_{AB} = T_A \circ T_B.$$

(This justifies our remarks in the paragraph following eqn. (1.2).)

In addition to being a complex vector space, the space $\mathbb{C}^n$ has another structure, namely that given by its 'inner product'. The **inner product** of two vectors in $\mathbb{C}^n$ is the complex number defined by

$$\langle(\xi_1, \cdots, \xi_n), (\eta_1, \cdots, \eta_n)\rangle = \sum_{i=1}^{n} \xi_i \overline{\eta_i} \ . \tag{1.5}$$

The rationale for consideration of this 'inner product' stems from the observation - which relies on basic facts from trigonometry - that if $x = (\xi_1, \xi_2), y = (\eta_1, \eta_2) \in \mathbb{R}^2$, and if one writes O, X and Y for the points in the plane with Cartesian co-ordinates $(0,0), (\xi_1, \xi_2)$ and (η_1, η_2) respectively, then one has the identity

$$\langle x, y\rangle = |OX|\ |OY|\ cos\ (angle\ XOY)\ ,$$

The point is that the inner product allows us to 'algebraically' describe distances and angles.

If $x \in \mathbb{C}^n$, it is customary to define

$$||x|| = (\langle x, x\rangle)^{\frac{1}{2}} \tag{1.6}$$

and to refer to $||x||$ as the **norm** of x. (In the notation of the previous example, we have $||x|| = |OX|$.)

One finds more generally (see [H], for instance) that the following relations hold for all $x, y \in \mathbb{C}^n$ and $\lambda \in \mathbb{C}$:

- $||x|| \geq 0$, and $||x|| = 0 \Leftrightarrow x = 0$
- $||\lambda x|| = |\lambda| \; ||x||$
- (**Cauchy-Schwarz inequality**)

$$|\langle x, y\rangle| \leq ||x|| \; ||y||$$

- (triangle inequality) $||x + y|| \leq ||x|| + ||y||$

More abstractly, one has the following definition:

DEFINITION 1.1 *A* **complex inner product space** *is a complex vector space, say V, which is equipped with an 'inner product'; i.e., for any two vectors $x, y \in V$, there is assigned a complex number - denoted by $\langle x, y\rangle$ and called the inner product of x and y; and this inner product is required to satisfy the following requirements, for all $x, y, x_1, x_2, y_1, y_2 \in V$ and $\lambda_1, \lambda_2, \mu_1, \mu_2 \in \mathbb{C}$:*
(a) (sesquilinearity) $\langle \sum_{i=1}^{2} \lambda_i x_i, \sum_{j=1}^{2} \mu_j y_j\rangle = \sum_{i,j=1}^{2} \lambda_i \overline{\mu_j} \langle x_i, y_j\rangle$
(b) (Hermitian symmetry) $\langle x, y\rangle = \overline{\langle y, x\rangle}$
(c) (Positive definiteness) $\langle x, y\rangle \geq 0$, *and* $\langle x, x\rangle = 0 \Leftrightarrow x = 0$.

The statement '$\mathbb{C}^n$ is the prototypical n-dimensional complex inner product space' is a crisper, albeit less precise version of the following fact (which may be found in basic texts such as [H], for instance):

PROPOSITION 1.2 *If V_1 and V_2 are n-dimensional vector spaces equipped with an inner product denoted by $\langle \cdot, \cdot\rangle_{V_1}$ and $\langle \cdot, \cdot\rangle_{V_2}$, then there exists a mapping $U : V_1 \to V_2$ satisfying:*

(a) U is a linear map (i.e., $U(\lambda x + y) = \lambda Ux + Uy$ for all $x, y \in V_1$); and

(b) $\langle Ux, Uy\rangle_{V_2} = \langle x, y\rangle_{V_1}$ for all $x, y \in V_1$.
Moreover, a such a mapping U is necessarily a 1-1 map of V_1 onto V_2, and the inverse mapping U^{-1} is necessarily also an inner product preserving linear mapping. A mapping such as U above is called a **unitary operator** *from V_1 to V_2.*

In particular, we may apply the above proposition with $V_1 = \mathbb{C}^n$ and any n-dimensional inner product space $V = V_2$. The following lemma and definition are fundamental. (We omit the proof which is not difficult and may be found in [H], for instance. The reader is urged to try and write down the proof of the implications $(i) \Leftrightarrow (ii)$.)

LEMMA 1.3 *Let V be an n-dimensional inner product space. The following conditions on a set $\{v_1, v_2, \cdots, v_n\}$ of vectors in V are equivalent:*

(i) there exists a unitary operator $U : \mathbb{C}^n \to V$ such that $v_i = Ue_i^{(n)}$ for all i.

(ii) $\langle v_i, v_j \rangle = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$.

The set $\{v_1, v_2, \cdots, v_n\}$ is said to be an **orthonormal basis** *for V if it satisfies the above conditions.*

If V is as above, and if $\{v_1, v_2, \cdots, v_n\}$ is any orthonormal basis for V, then it is easy to see that

(i) $v = \sum_{i=1}^n \langle v, v_i \rangle v_i$ for all $v \in V$; and

(ii) $\langle v, w \rangle = \sum_{i=1}^n \langle v, v_i \rangle \langle v_i, w \rangle$ for all $v, w \in V$.

Now if $T : V \to V$ is a linear transformation on V, the action of T may be encoded, with respect to the basis $\{v_i\}$, by the matrix $A \in M_{n \times n}(\mathbb{C})$ defined by

$$a_{ij} = \langle Tv_j, v_i \rangle \ .$$

We shall call A *the matrix representing* T *in the basis* $\{v_1, \cdots, v_n\}$.

It is natural to call an $n \times n$ matrix unitary if it represents a unitary operator $U : V \to V$ in some orthonormal basis; and it is not too difficult to show that a matrix is unitary if and only if its columns form an orthonormal basis for C^n.

More or less by definition, we see that if $A, B \in M_{n \times n}(\mathbb{C})$, the following conditions are equivalent:

(a) there exists a linear transformation $T : V \to V$ such that A and B represent T with repect to two orthonormal bases;

(b) there exists a unitary matrix U such that $B = UAU^{-1}$.

In (b) above, the U^{-1} denotes the unique matrix which serves as the *multiplicative inverse* of the matrix U. (Recall that the

multiplicative identity is given by the matrix I_n whose (ij)-th entry is δ_{ij} (defined in Lemma 1.3(ii) above); and that the matrix representing an operator is invertible if and only if that operator is invertible.)
Finally recall that the **trace** of a matrix $A \in M_n(\mathbb{C})$ is defined by [3]

$$Tr_n\ A \ = \ Tr\ A \ = \ \sum_{i=1}^{n}\ a_{ii}$$

and recall the following basic property of the trace:

PROPOSITION 1.4 *Suppose $A \in M_{m\times n}(\mathbb{C}), B \in M_{n\times m}(\mathbb{C})$. Then,*

$$Tr_m\ AB \ = \ Tr_n\ BA\ .$$

In particular, if $C, S \in M_n(\mathbb{C})$ and if S is invertible, then

$$Tr\ SCS^{-1} = Tr\ C\ ,$$

Proof: For the first identity, note that

$$Tr_m\ AB = \sum_{i=1}^{m}\left(\sum_{k=1}^{n} a_{ik}b_{ki}\right) = \sum_{k=1}^{n}\left(\sum_{i=1}^{m} b_{ki}a_{ik}\right) = Tr_n\ BA\ .$$

The second identity follows from the first, since

$$Tr\ SCS^{-1} = Tr\ CS^{-1}S = Tr\ CI_n = Tr\ C\ .$$

□

2 On commutators, numerical ranges and zero diagonals

We wish to discuss elementary proofs of the following three well-known results:

(A) A square complex matrix A has trace zero if and only if it is a commutator - i.e., $A = BC - CB$, for some B, C.

[3] Here and in the sequel, we shall write M_n instead of $M_{n\times n}$.

(B) If T is a linear operator on an inner product space V, then its numerical range $W(T) = \{\langle Tx, x\rangle : x \in V, ||x|| = 1\}$ is a convex set.

(C) A matrix $A \in M_n(\mathbb{C})$ has trace zero if and only if there exists a unitary matrix $U \in M_n(\mathbb{C})$ such that UAU^{-1} has all entries on its 'main diagonal' equal to zero.

As for the arrangement of the proof, we shall show that (C) follows from (B), which in turn is a consequence of the case $n = 2$ of (C). So as to be logically consistent, we shall first prove (C) when $n = 2$, then derive (B), then deduce (C) for general n, and finally deduce (A) from (C). Further, since the 'if' parts of both (A) and (C) are immediate (given the truth of Proposition 1.4), we shall only be concerned with the 'only if' parts of these statements.

Our proofs will not be totally self-contained; we will need one 'standard fact' from linear algebra. Thus, in the proof of Lemma 2.1 below, we shall need the fact that - at least in two-dimensions - every complex matrix has an 'upper triangular form'.

In the following proofs, we shall interchangeably think about elements of $M_n(\mathbb{C})$ as linear operators on $\mathbb{C}^n$ (or equivalently, on some n-dimensional complex inner product space with a distinguished orthonormal basis).

LEMMA 2.1 *If $A \in M_2(\mathbb{C})$ and $Tr\ A = 0$, then there exists a unitary matrix $U \in M_2(\mathbb{C})$ such that*

$$UAU^{-1} = \begin{pmatrix} 0 & * \\ * & 0 \end{pmatrix} .$$

Proof: To start with, we appeal to the fact - see [H], for instance - that every complex square matrix has an 'upper triangular form' with respect to a suitable orthonormal basis; in other words, there exists a unitary matrix $U_1 \in M_2(\mathbb{C})$ such that

$$U_1AU_1^{-1} = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} . \qquad (2.7)$$

Note - by Proposition 1.4 - that

$$a + c = Tr\ U_1AU_1^{-1} = Tr\ A = 0 ,$$

and so $c = -a$. In case $a = 0$, we may take $U = U_1$ and the proof will be complete.

So suppose $a \neq 0$. This hypothesis guarantees that the matrix A has the distinct 'eigenvalues' a and $-a$; i.e., we can find vectors x, y of norm 1 such that $U_1AU_1^{-1}x = ax$ and $U_1AU_1^{-1}y = -ay$. (In fact, $x = e_1^{(2)}$ and $y = pe_1^{(2)} + qe_2^{(2)}$ for suitable p and q with $q \neq 0$ (since $a \neq 0$). Thus x and y are lineary independent. Now, if $\alpha, \beta \in \mathbb{C}$, we have:

$$\begin{aligned}\langle U_1AU_1^{-1}(\alpha x + \beta y), (\alpha x + \beta y)\rangle &= a\langle(\alpha x - \beta y), (\alpha x + \beta y)\rangle \\ &= a(|\alpha|^2 - |\beta|^2 + 2iIm\ \alpha\bar{\beta}\langle x, y\rangle)\ .\end{aligned}$$

Now pick α, β to satisfy $|\alpha| = |\beta| = 1$ and $Im\ \alpha\bar{\beta}\langle x, y\rangle = 0$ - which is clearly possible. Independence of x and y and the fact that $\alpha, \beta \neq 0$ guarantee that $w = \alpha x + \beta y \neq 0$. Then, $\langle U_1AU_1^{-1}w, w\rangle = 0$.

Let $u_1 = \frac{w}{||w||}$, and let u_2 be a unit vector orthogonal to u_1. Let U_2 be the unitary operator on $\mathbb{C}^2$ such that $U_2^{-1}e_j^{(2)} = u_j$ for $j = 1, 2$. It is then seen that if $U = U_2U_1$ and $B = UAU^{-1}$, then

$$\begin{aligned}\langle Be_1^{(2)}, e_1^{(2)}\rangle &= \langle U_2(U_1AU_1^{-1})U_2^{-1}e_1^{(2)}, e_1^{(2)}\rangle \\ &= \langle (U_1AU_1^{-1})U_2^{-1}e_1^{(2)}, U_2^{-1}e_1^{(2)}\rangle \\ &= \langle (U_1AU_1^{-1})u_1, u_1\rangle \\ &= 0\ .\end{aligned}$$

Since $Tr\ B = Tr\ A = 0$, we conclude that the (2,2)-entry of B must also be zero; in other words, this U does the trick for us.

Proof of (B): It suffices to prove the result in the special case when V is two-dimensional. (*Reason:* Indeed, if x and y are unit vectors in V, and if V_0 is the subspace spanned by x and y, let T_0 denote the operator on V_0 induced by the matrix

$$\begin{pmatrix} \langle Tu_1, u_1\rangle & \langle Tu_2, u_1\rangle \\ \langle Tu_1, u_2\rangle & \langle Tu_2, u_2\rangle \end{pmatrix}\ ,$$

where $\{u_1, u_2\}$ is an orthonomal basis for V_0. The point is that T_0 is what is called a 'compression' of T and we have

$$\langle T_0x_0, y_0\rangle = \langle Tx_0, y_0\rangle \text{ whenever } x_0, y_0 \in V_0\ .$$

In particular, if we knew that $W(T_0)$ was convex, then the line joining $\langle Tx, x\rangle$ and $\langle Ty, y\rangle$ would be contained in the convex set $W(T_0)$ which in turn is contained in $W(T)$ (by the displayed inclusion above).)

Thus we may assume $V = \mathbb{C}^2$. Also, since $W(T - \lambda I_2) = W(T) - \lambda$ - as is readily checked - we may assume, without loss of generality that $Tr\ T = 0$. Then, by Lemma 2.1, the operator T is represented, with respect to a suitable orthonormal basis, by the matrix

$$\begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix} .$$

An easy computation then shows that

$$W(T) = \{ay\bar{x} + bx\bar{y} : x, y \in \mathbb{C}, |x|^2 + |y|^2 = 1\} .$$

Since $\{y\bar{x} : x, y \in \mathbb{C}, |x|^2 + |y|^2 = 1\} = \{z \in \mathbb{C} : |z| \leq \frac{1}{2}\}$, we thus find that

$$W(T) = \{az + b\bar{z} : z \in \mathbb{C}, |z| \leq \frac{1}{2}\}$$

and we may deduce the convexity of $W(T)$ from that of the disc $\{z \in \mathbb{C} : |z| \leq \frac{1}{2}\}$. □

Proof of (C): We prove this by induction, the case $n = 2$ being covered by Lemma 2.1.

So assume the result for $n - 1$, and suppose $A \in M_n(\mathbb{C})$. Then notice, by the now established (B), that

$$0 = \frac{1}{n}\sum_{i=1}^{n}\langle Ae_i^{(n)}, e_i^{(n)}\rangle \in W(A) .$$

Consequently, there exists a unit vector u_1 in $\mathbb{C}^n$ such that $\langle Au_1, u_1\rangle = 0$. Choose $u_2, \cdots .u_n$ be so that $\{u_1, \cdots, u_n\}$ is an orthonormal basis for $\mathbb{C}^n$, and let U be the unitary operator on $\mathbb{C}^n$ such that $U_1^{-1}e_i^{(n)} = u_i$ for $1 \leq i \leq n$. Then it is not hard to see that if $A_1 = U_1AU_1^{-1}$, then

- $\langle A_1e_1^{(n)}, e_1^{(n)}\rangle = 0$; and

- if B denotes the submatrix of A_1 determined by deleting its first row and first column, then, $Tr_{n-1}\ B = Tr_n\ A_1 = Tr_n\ A = 0$; and hence by our induction hypothesis, we can choose an orthonormal basis $\{v_2, \cdots, v_n\}$ for the subspace spanned by $\{e_2^{(n)}, \cdots, e_n^{(n)}\}$ such that $\langle Bv_j, v_j \rangle = 0$ for all $2 \leq j \leq n$.

We then find that $\{u'_1 = u_1, u'_2 = U^{-1}v_2, \cdots, u'_n = U^{-1}v_n\}$ is an orthonormal basis for $\mathbb{C}^n$ such that $\langle Au'_i, u'_i \rangle = 0$ for $1 \leq i \leq n$. Finally, if we let U be a unitary matrix so that $U^{-1}e_i^{(n)} = u'_i$ for each i, then UAU^{-1} is seen to satisfy

$$\langle UAU^{-1}e_i^{(n)}, e_i^{(n)} \rangle = 0 \quad \text{for} \quad \text{all} \quad i\ .$$

Proof of (A): By replacing A by UAU^{-1} for a suitable unitary matrix U, we may, by (C), assume that $a_{ii} = 0$ for all i. Let $b_1, b_2, \cdots, b_n$ be any set of n distinct complex numbers, and define

$$b_{ij} = \delta_{ij} b_j\ ,\ c_{ij} = \begin{cases} 0 & \text{if } i = j \\ \frac{a_{ij}}{b_i - b_j} & \text{if } i \neq j \end{cases}\ .$$

It is then seen that indeed $A = BC - CB$.

3 Extensions

It is natural to ask if complex numbers have anything to do with the result that we have called (A). The reference [AM] extends the result to more general fields.

In another direction, one can seek 'good infinite-dimensional analogues' of (A); one possible such line of generalisation is pursued in [BP], where it is shown that 'a bounded operator on Hilbert space is a commutator (of such operators) if and only if it is not a compact perturbation of a non-zero scalar'.

References:

[AM] Albert, A.A., and Muckenhoupt, B., *On matrices of trace zero*, Michigan Math. J., **4**, (1957), 1-3.

[BP] Brown, A., and Pearcy, C., *Structure of Commutators of operators*, Ann. of Math., **82**, (1965), 112-127.

[H] Halmos, P.R., *Finite-dimensional vector spaces*, Van Nostrand, London, 1958.

Address for correspondence:
The Institute of Mathematical Sciences
Chennai.