A GENERALIZATION OF THE CAYLEY-HAMILTON THEOREM

by

Azaria Paz

Technical Report #306

November 1983
Let $L = \{ m_1 \}$, $|L| \geq 2$, be a finite set of square $n \times n$ matrices. Let $L_p$ be the linear space with rank $r_p (\leq n^2)$ spanned by the set of matrices $\{ m : m \in L^*, |m| \leq p \}$ and let $c(L)$ be defined by $c(L) = \min \{ p : r_p = r_{p+1} \}$, where $|X|$ denotes the number of elements in a set $X$ or the length of a word $X$, and $L^*$ denotes the set of all finite words over $L$ (here a word is a sequence of matrices from $L$ and is understood as the matrix product of the matrices in the sequence). It is shown in the paper that $c(L) \leq \left\lfloor \frac{n^2+2}{3} \right\rfloor$. The proof of the above bound is based on combinatorial arguments and it is not known whether the bound is sharp. It follows from the Cayley-Hamilton Theorem that $c(L) \leq n-1$ if $L$ is a singleton.

It is also shown (a result of H.W. Leustra Jr.) that the bound $c(L) \leq n-1$ holds for commutative matrices.

---

* This work has been supported in part by NSF Grant No.MCS 8012716. This report is an extended version of a paper with the same title to appear in Linear and Multilinear Algebra.

An abstract of a preliminary version of this paper has been presented at the Ecole de Printemps d'Informatique Theorique, "Serie formelles en variables non-commutatives et applications", Vieux-Boucau les Bain (Landes), France, 23-27 Mai, 1977.
1. INTRODUCTION

Let $\Sigma = \{A_0, A_1, \ldots, A_t\}$ be an alphabet consisting of $n \times n$ matrices $A_i$ over a commutative field. For the sake of simplicity we assume in the sequel that there are only two matrices in $\Sigma$, $A_0$ and $A_1$, but this restriction is not essential.

We shall use the following notations: $\Sigma^k$ denotes the set of all products of length $k$ and $\Sigma^*$ denotes the set of all products of finite length over $\Sigma$ (including the unit matrix $I$); $L_1$ denotes the linear space spanned by $I \cup \Sigma \cup \ldots \cup L_1$ (the generators of the space are the product matrices in this set) and $L_\ast$ denotes the linear space spanned by $\Sigma^*$; $r_1$ and $r_\ast$ denote the rank of $L_1$ and $L_\ast$ respectively.

It is easy to show that there exists an integer $i$ such that

$$r_1 = r_\ast \leq n^2.$$

Let $c(\Sigma) = \min\{i: r_1 = r_\ast\}$. It follows trivially from the Cayley-Hamilton theorem that if $\Sigma$ consists of a single $n \times n$ matrix then, for that $\Sigma$, $c(\Sigma)$ is bounded by $n-1$ and this bound is sharp. To the best knowledge of this author, the natural problem of (sharply) bounding $c(\Sigma)$ in the general case has not been considered in the literature so far. Motivated by a possible application to the mechanics of isotropic continua, Spencer & Rivlin [7],[8] have studied a related problem for the restricted case where the matrices in $\Sigma$ are $3 \times 3$. They have shown, among other things, that for this case $c(\Sigma) \leq 5$. It is interesting to note that for the general case ($n \times n$ matrices), if the matrices in $\Sigma$ are pairwise permutable then $c(\Sigma) \leq n-1$ as in the single matrix case. This result is due to H.W. Lenstra from the University of Amsterdam (private communication) and is reproduced in this paper with his kind permission (see Section 5). Our main result here is that $c(\Sigma) \leq \left\lfloor \frac{n^2+2}{3} \right\rfloor$, where the method of proof is based mainly on combinatorial arguments. For the case where $n = 3$,
this bound reduces to \( c(\Sigma) \leq 4 \) which is better than Spencer & Rivlin's bound mentioned above. J.W. Carlyle from U.C.L.A. tried to construct, with the help of a computer program, sets of matrices which will attain our bound. He managed to show (private communication) that the bound is sharp for \( n \leq 4 \). On the other hand, based on those experiments, we believe now that our bound is not sharp for \( n > 4 \). In fact we would like to suggest the conjecture that in the general case (\( n \times n \) matrices) \( c(\Sigma) \leq 2(n-1) \) (which is compatible with \( \left\lfloor \frac{n^2 + 2}{3} \right\rfloor \) for \( n \leq 4 \)). We cannot prove this conjecture and we will leave it therefore as an open problem. The problem considered in this paper is somewhat related to the invariant Theory of Matrices (see e.g. [3],[4],[5]).

As a corollary to our main theorem we have the result that every matrix polynomial over a set of \( n \times n \) matrices \( \Sigma \) can be expressed as a polynomial over \( \Sigma \) of degree \( \leq \left\lfloor \frac{n^2 + 2}{3} \right\rfloor \).

2. EXAMPLES

For \( n=2 \) the bound is attained by the matrices

\[
A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}; \quad B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}; \quad \Sigma = \langle A, B \rangle.
\]

The set of generators of \( \Sigma^* \) is

\[ \{ I, A, B, AB \} \] with \( c(\Sigma) = 2 = \left\lfloor \frac{4+2}{3} \right\rfloor \)

and \( r^* = 4 = n^2 \).

For \( n=3 \) (J.W. Carlyle)

\[
A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}; \quad B = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & -1 \\ 1 & 1 & 1 \end{bmatrix};
\]

The set of generators is

\[ \{ I, A, B, AB, BA, BB, ABB, BAB, BABB \} \]

\( c(\Sigma) = 4 = \left\lfloor \frac{9+2}{3} \right\rfloor \) and \( r^* = 9 = n^2 \).
The next example is also due to J.W. Carlyle.

For \( n = 4 \) the bound is attained by the following set of matrices:

\[
A = \begin{bmatrix} 1 & 2 & 2 & 2 \\ -1 & 0 & 0 & -2 \\ 0 & 0 & 2 & 0 \\ 1 & 2 & 1 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad \Sigma = \langle A, B \rangle .
\]

The set of generators of \( \Sigma^* \) is:

\[
\{ I, A, B, A^2, AB, BA, A^3, A^2B, ABA, BA^2, A^3B, A^2BA, ABA^2, A^3BA, A^2BA^2, A^3BA^2 \}
\]

with \( c(\Sigma) = 6 = \left\lceil \frac{16+2}{3} \right\rceil \) and \( r_\ast = 16 = n^2 \).

The above examples show that the bound is sharp for \( n \leq 4 \).

Notice e.g. that for \( n = 4 \) if \( \Sigma \) contains a single matrix, then \( c(\Sigma) \) would be equal to 3 (by the Cayley-Hamilton theorem) showing that the bound derived from Cayley-Hamilton's theorem does not hold for the general case.

On the other hand, we have that

\[
\left\lfloor \frac{n^2+2}{3} \right\rfloor = 2(n-1) \quad \text{for} \quad 2 \leq n \leq 4,
\]

and

\[
\left\lfloor \frac{n^2+2}{3} \right\rfloor > 2(n-1) \quad \text{if} \quad n > 4.
\]

No example has been found yet showing that the bound is sharp for \( n > 4 \).

The following example, for \( n = 5 \), which is also due to J.W. Carlyle has

\[
c(\Sigma) = 8 = 2(n-1) < 9 = \left\lceil \frac{n^2+2}{3} \right\rceil .
\]

\[
A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 0 & 0 & 0 & -1 \\ -1 & 1 & 1 & 1 & 0 \\ 0 & 1 & -1 & -1 & 0 \\ 1 & -1 & 0 & -1 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix} .
\]

The set of generators is:

\[
\]

with \( c(\Sigma) = 8 \) and \( r = 25 = n^2 \).
3. SOME ADDITIONAL NOTATIONS AND USEFUL FACTS

Let the matrices in $\Sigma$ be $n \times n$.

Let $m = n + t > n$. In this and in the following section we assume $|\Sigma| = 2$ and denote the matrix $A_i A_i^t \ldots A_i^{m}$ by word $i_1 i_2 \ldots i_m = u$. As $i_j \in \{0,1\}$ the word $u$ is also a binary number which will be denoted by $\bar{u}$. The matrix $u$ which belongs to $L_m$ will be termed reducible if it belongs to $L_{m'}$ for some $m' < m$. There are several matrices which are trivially reducible, e.g. the matrices denoted by words $\begin{array}{c}
\bar{m} \\bar{m} \\
00 \ldots 0 \\
00 \ldots 01 \\
01 \ldots 1 \\
100 \ldots 0
\end{array}$, $\begin{array}{c}
\bar{m} \\bar{m} \\
11 \ldots 10 \\
11 \ldots 11 \\
10 \ldots 10
\end{array}$ (As $m > n$ the Cayley-Hamilton theorem applies to subwords of the above words.) Let $w$ be the word consisting of alternating zeros and ones and starting with zero and let $w'$ be the word of alternating zeros and ones and starting with one. For a given word $u = i_1 i_2' \ldots i_m'$, the subword $i_j+1 i_j+2' \ldots i_j+k$ is denoted by $u_{j,k}$ (thus $u = u_{0,m}$ and $u_j = u_{j-1,j}$).

The following notations for matrices in $L_m$ will be used in the next section:

$u \sim v$ for $u$ is linearly dependent on $v$ (and vice versa), modulo $L_{m-1}$;

$u \propto (v, w)$ for $u$ is a linear combination of $v$ and $w$, modulo $L_{m-1}$;

$u_{ij} + (u_{k}, \ell, u_{s}, t)$ |$\sim (u', u'')$ for substituting in $u$, $u_{k, \ell}$ for $u_{i, j}$ and $u_{s, t}$ for $u_{i, j}$ results in $u'$ and $u''$ correspondingly.
We need also the following facts:

1. A word is reducible if it has a reducible subword.

2. Every word (of length $m$) which is not trivially reducible and differs from $w$ and $w'$ (see definition above) has at least 3 different subwords of length $k$ for $2 \leq k \leq m-2$ (if a window of length $k$ is moved over the word and it assumes only two different digit patterns then the patterns are either alternating, as for the words $w$ and $w'$, or they change only once, at the first or last move, correspondingly to trivially reducible words.

4. A COMBINATORIAL LEMMA

The proof of our main theorem (next section) depends on the following:

Lemma 1: Let $m = n+t$, $t \geq 1$. If any of the following $m-1$ conditions

\begin{align*}
    a_k: & \quad r_k - r_{k-1} \leq 2 \quad (2 \leq k \leq m-2) \\
    a_{m-1}: & \quad r_{m-1} - r_{m-2} \leq 1
\end{align*}

holds true then $c(\Sigma) \leq m-1$.

Proof:

Divide the words in $\Sigma^m$ except $w$ and $w'$ into four disjoint sets as follows:

- Set $I = \{ u: 0 \leq \tilde{u} < \tilde{w} \}$
- Set $I' = \{ u: \tilde{w} < \tilde{u} < 11...1 \}$
- Set $II = \{ u: \tilde{w} < \tilde{u} < 01...1 \}$
- Set $II' = \{ u: 100...0 \leq \tilde{u} < \tilde{w}' \}$

i.e., if the words in $\Sigma^m$ are ordered in a list according to the numerical value $\tilde{u}$ then the list is subdivided as in the following table:
We will show that, if any of the conditions of the lemma holds for the given $m > n$, then all the words in $E^m$ are reducible, thus showing that $c(E) \leq m-1$. This will be done for each of the four sets $I, I', II, II'$ separately, and also for the two words $w$ and $w'$ in particular.

Assume first that one of the conditions $a_k$ for $2 \leq k \leq m-2$ holds. Let $u_1, u_2, u_3$ be three different words of length $k$. It follows from the definition of $a_k$ that either $u_i \sim u_j$ or $u_s \approx (u_i, u_j)$ where $i, j, s \in \{1, 2, 3\}$, $i \neq j \neq s, i \neq s$. 
The following cases must be considered:

1. \( u \in I \). If \( u \) is not trivially reducible then \( u \) has at least 3 different subwords of length \( k \), say \( u_1, u_2, u_3 \). Assume that \( u_1 > u_2 > u_3 \).

If \( u_2 > u_3 \) then \( u_2 + u_3 \nmid u' \) with \( \tilde{u}' < \tilde{u} \) and \( u' \sim u \). Else \((u_1 \preceq (u_2,u_3))\), \( u_1 + (u_2,u_3) \nmid (u',u'') \) with \( \tilde{u}'', \tilde{u}'' < \tilde{u} \), and \( u \preceq (u',u'') \). The reduction can be repeated for \( u' \) and \( u'' \), and so on until the words \( \text{000...0} \) and \( \text{000...01} \) are reached and those words are trivially reducible. One can handle the set \( I' \) in a similar way.

2. If \( u = w = \text{0101...} \) then let \( v = w_{1,n+1} = \underbrace{1010...}_n \). Thus \( v \) is one of the terms in the expansion of \( (0+1)^n \) (0 and 1 denote matrices).

One can use the ordinary Cayley-Hamilton Theorem for the matrix \( 0 + 1 \) in order to show that \( v \) can be expressed as a linear combination of other words of length \( n \) (modulo words of length \( n-1 \) or less). Thus \( w \) can be expressed as a linear combination of other words of length \( m \) (modulo words of length \( m-1 \) or less) all starting with \( 0 \) (the substitution of \( v \) in \( w \) does not affect the first letter of \( w \)) and therefore different from \( w' \). One can handle the word \( w' \) in a similar way, expressing it as a combination of words of length \( m \) (modulo words of length \( m-1 \) or less) all different from \( w \).

3. Assume that \( u \in \mathcal{I} \). By definition \( \tilde{u} \succ \tilde{w} \), and \( u \leq 0111... \)

Let \( k \) be the index of the condition \( a_k \) in the lemma, and \( 2 \leq k \leq m-2 \), assumed to hold true. We shall consider the following subcases:

3.1 \( u_{o,k+1} = w_{o,k+1} \). This implies \( (\tilde{u} > \tilde{w}) \) that \( \tilde{u}_{i,i+k} > \tilde{w}_{i,k+k} \) for some \( i \). Thus, \( u_{o,k} = \text{0101...} ; u_{1,k+1} = \text{1010...} \) and
$u_{i,i+k}$ differs from both $u_{o,k}$ and $u_{1,k+1}$. Either $u_{o,k} \sim u_{i,i+k}$ or $u_{1,k+1} \sim (u_{o,k} u_{i,i+k})$. In the first case

\[ u_{o,k} \sim u_{i,i+k} \sim u' \] where if $u_{i,i+k}$ starts with a zero then $u'$ belongs to case 3.2 below (recall that $\tilde{u}_{i,i+k} > \tilde{w}_{i,i+k}$) while if $u_{i,i+k}$ starts with a one, then $u' \in I'$ (for the same reason) and can be handled as in case 1.

In the second case

\[ u_{1,k+1} \sim (u_{o,k}, u_{i,i+k}) \sim (u', u'') \] where $u'$ starts with a double zero so that it belongs to $I$ and can be handled as above. If $u_{i,i+k}$ starts with a zero then $u''$ too belongs to $I$ (it starts with a double zero) and can be handled accordingly. Otherwise $\tilde{u}_{o,k+1} > \tilde{w}_{o,k+1}$ implying that $\tilde{u}_{i,k+1} > \tilde{w}_{i,k+1}$ (the first digit in both words is 0 as $u \in II$) and this implies that

11 is a subword of $u_{i,k+1}$.

3.2 $\tilde{u}_{o,k+1} > \tilde{w}_{o,k+1}$ implying that $\tilde{u}_{i,k+1} > \tilde{w}_{i,k+1}$ (the first digit in both words is 0 as $u \in II$) and this implies that 11 is a subword of $u_{i,k+1}$.

3.2.1 $u_{o,3} = 010$ and $k = 2$. The words 10, 01 and 11 are subwords of $u$ (as mentioned above). Also $u_{o,4} = 0101$, ($u \in II$).

If 10 $\sim$ 11 then $u_{13} \sim 11 \sim 0111...$ and this case will be discussed below (3.2.3). Otherwise, 01 $\sim$ (10, 11) and therefore $(01=\tilde{u}_{2,4})u_{2,4} \sim (10, 11) \sim (0111..., 0110...)$ which cases are discussed below (3.2.2 and 3.2.3).

3.2.2 $u_{o,4} = 0110$ and $k = 2$. If 01 $\sim$ 10 then the two ones can be moved to the beginning of the word resulting in a word $u' \in I'$ which can be handled according to case 1. Otherwise, 11 $\sim$ (01, 10) inducing $u_{1,5} = (01, 10) \sim (u', u'')$ where $u' = 0010...$ and $u'' = 0100...$. Both $u'$ and $u''$ are in $I$ and can be reduced according to case 1.
3.2.3 $u_{0,4} = 0111$ and $k = 2$. If $u$ is not trivially reducible then it contains the subword 10. Either $10 \sim 11$ in which case the substitution $10 \rightarrow 11$, when applied successively will reduce $u$ into the word $011\ldots 1$ which is trivially reducible, or $01 \approx (10,11)$ inducing the reduction:

$$u_{0,2} + (10,11) \triangleright (u',u'')$$

where $u' = 1011\ldots$ and $u'' = 1111\ldots$ Both words are in $I'$ and can be reduced as in case 1.

3.2.4 Same as 3.2.1. ($u_{0,3} = 010$) but $k \geq 3$. Then $u_{1,k+1} > u_{2,k+2}$ (trivial) and $u_{2,k+2} > u_{0,k}$ as first occurrence of the subword 11 which must exist in $u_{2,k+2}$ is closer to the left in $u_{2,k+2}$ than in $u_{0,k}$ (if it is also a subword of $u_{0,k}$). Thus, either $u_{0,k} \sim u_{2,k+2}$ or $u_{2,k+1} = (u_{0,k},u_{2,k+2})$.

In the second case we can apply the reduction $u_{1,k+1} + (u_{0,k},u_{2,k+2}) \triangleright (u',u'')$ where both $u'$ and $u''$ start with a double zero (both $u_{0,k}$ and $u_{2,k+2}$ start with zero), are in $I$ and can be reduced as in case 1.

In the first case we can apply (successively if needed) the reduction $u_{0,k} + u_{2,k+2} \triangleright u'$ until a word of the form $0110\ldots$ is reached (case 3.2.5 below) or a word of the form $0111\ldots$ is reached (case 3.2.6 below).

Remark: Notice that if $u = u_{0,5}$ (i.e. $u$ is of length 5) then $k = 3$ (we assumed that $k \geq 3$) necessarily and $u$ must be equal to $01011$ ($\bar{u} > \bar{w}$). In this case the second reduction $u_{0,k} + u_{2,k+2} \triangleright u'$ will result in $u' = 01111$ which is trivially reducible.

3.2.5 Same as 3.2.2 ($u_{0,4} = 0110$) but $k \geq 3$. Assume first that $u$ is of length 5. Then $u = 0110x$, $x \in \{0,1\}$ and $k = 3$. If $011 \sim 110$ then $u_{0,3} + u_{1,4} \triangleright 100x \in I'$ and proceed as in
case 1. Otherwise $10x = (011,110)$ so that the following reduction applies: $u_{2,5} \rightarrow (u_0,3,u_1,4) \rightarrow (u',u'')$ where $u' = 01011$ which can be reduced as in the remark following the previous case, and $u'' = 01110$ which is the case 3.2.6 below. Assume now that $u$ has length $\geq 6$, i.e. $u = 0110xy...$, $x,y \in \{0,1\}$. If $0xy \neq 011$ then either $u_{3,3+k} \sim u_{0,k}$ in which case $u_{3,k+3} \rightarrow 011011...$ or $u_{1,k+1} \rightarrow (u_{3,3+k},u_{0,k})$ in which case $u_{1,k+1} \rightarrow (u_{3,3+k},u_{0,k}) \rightarrow (u',u'')$ where both $u'$ and $u''$ start with a double zero and can be reduced as in case 1.

We can assume therefore that $u = 011011...$ Then either $u_{1,k+1} \sim u_{2,k+2}$ leading to the reduction $u_{2,k+2} \rightarrow u_{1,k+1} \rightarrow 011101...$ which is the case 3.2.6 below or $u_{0,k} \rightarrow (u_{2,k+2},u_{1,k+1})\rightarrow (u',u'')$ where $u' = 10101... (if k = 3)$ or $u' = 1011... (if k > 3)$ and $u'' = 110...$ so that both $u'$ and $u''$ are in $I'$ and can be reduced as in case 1.

3.2.6 $u_{0,4} = 0111$. If $u$ has no additional zeros then $u$ is trivially reducible. Otherwise, let $u = 0111...10...$ i.e. after the first zero there are $s \geq 3$ ones. Then either $u_{s+1-k,s+1} \sim u_{s+2-k,s+2}$ in which case successive application of the reduction $u_{s+2-k,s+2} \rightarrow u_{s+1-k,s+1}$ will increase the length of the sequence of ones until the trivially reducible word $01...1$ is reached, or $u_{0,k} \rightarrow (u_{s+1-k,s+1},u_{s+2-k,s+2})$ which induces the reduction: $u_{0,k} \rightarrow (u_{s+1-k,s+1},u_{s+2-k,s+2}) \rightarrow (u',u'')$ where both $u'$ and $u''$ start with a double one $(k \geq 3)$, are in $I'$ and can be reduced as in case 1.
4. The case left is the case where \( k = m - 1 \). The corresponding condition in the lemma is that \( r_{m-1} - r_{m-2} \leq 1 \), i.e., every two words of length \( m - 1 \) both of them not trivially reducible are linearly dependent (modulo words of length \( \leq m - 2 \)) one on the other. Then, either one of the words \( u_{0,m-1} \) and \( u_{1,m} \) is reducible and we are done, or \( u_{0,m-1} = u_{1,m} \) in which case all the digits of \( u \) are equal to one another so that \( u \) is trivially reducible, or \( u_{0,m-1} \) can be substituted for \( u_{1,m} \) in \( u \) to get a new word \( u' \), \( u = u' \), starting with two equal digits. This reduction can be repeated until a word with \( m - 1 \) equal digits is reached, a word which is trivially reducible. The proof is now complete.

5. MAIN THEOREM

The following facts are easily proved (the notations are defined in the introduction):

1. For all integers \( i \), \( r_i \leq r_{i+1} \)
2. If for some \( i \), \( r_i = r_{i+1} \) then \( r_{i+1} = r_{i+2} = r_* \).
3. It follows from 2. above that there is a \( k \) such that
   \[ 1 = r_0 < \ldots < r_k = r_{k+1} = r_* \leq n^2 \]
   this implies that \( k+1 \leq r_k \leq n^2 \) or \( k \leq n^2 - 1 \).

Let \( |\Sigma| \) denote the number of matrices in \( \Sigma \), and let

We want to prove the following:
Theorem 1. If \( |\Sigma| = 2 \) and \( \sigma(\Sigma) = n \geq 2 \) then \( c(\Sigma) \leq \left\lfloor \frac{n^2+2}{3} \right\rfloor \).

Corollary: Every matrix polynomial over \( \Sigma \) with \( \sigma(\Sigma) = n \) can be expressed as a polynomial over \( \Sigma \) of degree \( \leq \left\lfloor \frac{n^2+2}{3} \right\rfloor \).

Proof: Theorem 1 is implied by lemma 1 as follows:

Assume that \( c(\Sigma) = m = n+t \), \( t \geq 1 \); \( \sigma(\Sigma) = n \). Then \( r_0 = 1 \) and \( r_1 \geq 3 \) (otherwise \( c(\Sigma) \leq n-1 \) by the Cayley-Hamilton theorem). Our assumption that \( c(\Sigma) = m > m-1 \), contrary to the consequence of the lemma, implies that its premise is false so that none of the conditions \( a_k \) holds: i.e., \( r_i - r_{i-1} \geq 3 \) for \( 2 \leq i \leq n+t-2 \) and \( r_{m-1} - r_{m-2} \geq 2 \). Thus,

\[
\begin{align*}
n^2 & \geq r_m = r_{m-2} + (r_{m-1} - r_{m-2}) + (r_{m} - r_{m-1}) \\
& \text{with} \\
r_{m-2} &= r_1 + (r_2 - r_1) + \ldots + (r_{m-2} - r_{m-3}) \geq 3(m-2),
\end{align*}
\]

\( r_{m-1} - r_{m-2} \geq 2 \) and \( r_m - r_{m-1} \geq 1 \), (otherwise \( c(\Sigma) = m-1 \)) resulting in \( n^2 \geq 3(m-2) + 3 = 3(m-1) \). If 3 is not a factor of \( n \) then the above inequality implies that \( n^2 \geq 3(m-1) + 1 \) or \( \left\lfloor \frac{n^2-1}{3} \right\! \right\rfloor \geq m-1 \). If 3 is a factor of \( n \) then \( \left\lfloor \frac{n^2}{3} \right\rfloor \geq m-1 \) implies that \( \left\lfloor \frac{n^2-1}{3} \right\! \right\rfloor \geq m-1 \). In both cases we have that \( \left\lfloor \frac{n^2-1}{3} \right\! \right\rfloor \geq m-1 \) or \( m \leq \left\lfloor \frac{n^2+2}{3} \right\rfloor \).

Q.E.D.

Notice that the above proof does not depend on the value of \( |\Sigma| \). If lemma 1 is true for \( |\Sigma| \geq 2 \), which it is (see remark 1 after the last section), then so is Theorem 1.
6. THE COMMUTATIVE CASE

The following theorem is due to H.W. Lenstra, Jr. from the Mathematical
Institute of the University of Amsterdam (private communication).

Theorem 2: If all the matrices \((n \times n)\) in \(\Sigma\) are pairwise permutable then
\(c(\Sigma) \leq n-1\).

Proof:

Let \(\Sigma = \{A_1, \ldots, A_t\}\) over the complex number field. It follows from
known theorems [2, Chap. VII, §3, Corol. 2, and Chap. VI; §6.3] that, under the
above conditions, there exists a non-singular matrix \(U_1\) s.t. the matrices
\(UA_1U^{-1} = A'_1\) are quasi-diagonal, having the same block structure, as follows:
\(A'_1 = \{A'_{11}, \ldots, A'_{1k}\}\) where \(A'_{jk}\) is an upper triangular matrix of order \(n_j\)
(the parameters \(k\) and \(n_j\) are fixed, not depending on \(i\), and \(\sum n_j = n\))
and all the entries in the main diagonal of \(A'_{ij}\) are equal. (A matrix is
quasi-diagonal if all its entries are zero except for the diagonal square blocks
listed inside the \{\} brackets.)

Now let \(\Sigma' = \{A'_1, \ldots, A'_t\}\), let \(B = B_1 \ldots B_n \in \Sigma'^n\), and let \(\lambda_{ij}\) be
the (constant) value of the entries in the diagonal of the \(j\)-th block of \(B_i\).
We have that
\[
\prod_{i=1}^{n_1} (B_i - \lambda_{i1} I) \prod_{i=n_1+1}^{n_1+n_2} (B_i - \lambda_{i2} I) \ldots \prod_{i=n-n_k+1}^{n} (B_i - \lambda_{ik} I) = 0.
\]
(The \(j\)-th factor in the above product has all the entries in the \(j\)-th diagonal
block equal to zero.)

It follows that the matrix \(B\) can be expressed as a linear combination
of products of up to \(n-1\) matrices in \(\Sigma'\):
\[ B = B_1 \ldots B_n = \alpha_1 C_1 + \alpha_2 C_2 \ldots + \alpha_t C_t, \]

where the \( C_i \) are words over \( \Sigma^S \) with \( 0 \leq s \leq n-1 \).

This completes the proof.

Notice, however, that there exists a commutative semigroup of \( n \times n \) matrices that contains at least \( \left\lfloor \frac{n}{2} \right\rfloor \times \left\lfloor \frac{n}{2} \right\rfloor + 1 \) linearly independent matrices (the semigroup containing all the matrices having a single entry equal to 1 in the upper \( \left\lfloor \frac{n}{2} \right\rfloor \times \left\lfloor \frac{n}{2} \right\rfloor \) right corner, with all the other entries equal to 0, and containing also the identity matrix).

7. REMARKS

1. In proving Lemma 1 we assumed that \( |\Sigma| = 2 \). This assumption is not necessary, the lemma being true for the general case with \( |\Sigma| > 2 \). This can be shown as follows:

Let \( \Sigma = \{A_0, \ldots, A_i, \ldots, A_{t-1}\} \) with \( |\Sigma| = t > 2 \). Let \( u \) denote a word in \( \Sigma^m, \ m > n \). Denote by \( u^{(ij)} \) the word of length \( m \),
\[ u^{(ij)} = ijjj \ldots, i \neq j \text{ and } 0 \leq i, j \leq t-1. \]

If \( u \) is not trivially reducible and \( u \neq u^{(ij)} \) then \( u \) has at least 3 subwords of length \( k \), \( 2 \leq k \leq m-2 \). One can therefore, for any fixed \( k \), corresponding to one of the conditions of Lemma 2, reduce \( u \) to trivially reducible words or to words of the form \( u^{(ij)} \). After the first reduction, every word \( u^{(ij)} \) can be reduced to trivially reducible words by a procedure similar to the procedure used in the proof of Lemma 1, keeping \( i \) and \( j \) fixed all along the reduction process. The case where \( k = m-1 \) carries over too, as the proof for this case (case 4 in the proof of the lemma) is independent of the number of elements in \( \Sigma \). It follows that Theorem 1 is true for \( |\Sigma| \geq 2 \).
2. While it is clear that the semigroup of $n \times n$ matrices generated by a single matrix has at most $n-1$ linearly independent matrices (a trivial consequence of the Cayley-Hamilton Theorem), this is certainly not true for general semigroup of matrices.

It is obvious that there are semigroups of matrices containing $n^2$ linearly independent matrices (see also the theorem of Burnside [9, p. 194] which guarantees the existence of nontrivial such semigroups). On the other hand, if a given semigroup of matrices contains $f(n) < n^2$ linearly independent matrices then the bound of our theorem can be reduced to

$$c(\Sigma) \leq \left\lceil \frac{f(n) + 2}{3} \right\rceil$$

as is easily seen.

3. Notice that Lemma 1 is true for any algebra over a field in which each element satisfies a polynomial of degree $\leq n$, and Theorem 1 is true if in addition, the algebra has dimension $\leq n^2$.

4. Prof. J.W. Carlyle from U.C.L.A. tried to construct, with the aid of a computer program, semigroups of matrices which will prove that the bound given in our theorem 1 here is sharp. Based on those experiments we would like to suggest the following conjecture:

Let $m = 2n-1$. If any of the following $m-2$ conditions

$$r_k - r_{k-1} \leq k \quad \text{for} \quad 1 \leq k \leq n-1$$

$$r_k - r_{k-1} \leq 2n-k-2 \quad \text{for} \quad n \leq k \leq 2n-2$$

holds true then $c(\Sigma) \leq m-1$ (for $|\Sigma| = 2$ containing $n \times n$ matrices).
If the above conjecture is true then we would have that $c(\Sigma) \leq 2(n-1)$ in general. This can be shown as follows: if $c(\Sigma) = m \geq 2n-1$ then none of the conditions in the conjecture can hold or $r_0 = 1$, $r_1 - r_0 \geq 2$, $r_2 - r_1 \geq 3 \ldots r_{n-1} - r_{n-2} \geq n$, $r_n - r_{n-1} \geq n-1 \ldots r_{2n-2} - r_{2n-1} \geq 1$. Thus $r_{2n-2} \geq 1+2 + \ldots + n-1 + n + n-1 + \ldots + 1 = 2 \frac{n(n-1)}{2} + n = n^2 > r_n$. So $c(\Sigma) \leq 2n-2$ necessarily. We do not have any counter example to disprove the above conjecture. Notice that all the examples in Section 2 satisfy the conditions of our conjecture.

5. A referee to which the paper has been sent remarked that the bound $c(\Sigma) \leq 2(n-1)$ is attained for $\Sigma = <A,B>$ where

$$A = \begin{bmatrix} 1 & 0 \\ w & 0 \\ w^2 & 0 \\ 0 & 1 \\ w^{-1} & \end{bmatrix} \quad ; \quad B = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

and $w$ is a primitive $n$-th root of unity.

ACKNOWLEDGEMENT

The author is indebted to the referees for many helpful comments.
BIBLIOGRAPHY


