Problem 5. Two n by n matrices M, N with real entries are called strongly independent if the matrix $aI + bM + cN$ is invertible for any real numbers a, b, c which are not all 0.

a) Show that if $MN = NM$ then M, N are not strongly independent.

b) For any n which is a multiple of 4 construct two strongly independent matrices of size n.

(Here I denotes the identity matrix.)

Solution. Let us start by recalling some basic facts from linear algebra. If V is a finite dimensional vector space over the field $\mathbb{C}$ of complex numbers and $T : V \longrightarrow V$ is a linear transformation then there exists a $\lambda \in \mathbb{C}$ and a non-zero vector $v \in V$ such that $Tv = \lambda v$ (we say that λ is an eigenvalue for T and v is an eigenvector corresponding to the eigenvalue λ). Furthermore, after choosing a basis of V, we may represent T by an $n \times n$ matrix D (where n is the dimension of V) and λ is an eigenvalue of T if and only if the matrix $\lambda I - D$ is not invertible, which is equivalent to the condition $\det(\lambda I - D) = 0$. The determinant $\det(\lambda I - D)$ is a polynomial function of λ of degree n and therefore λ is an eigenvalue of T if and only if it is a root of the polynomial $\det(\lambda I - D)$ (called the characteristic polynomial of T). So the existence of an eigenvalue is an immediate consequence of the Fundamental Theorem of Algebra, which asserts that any non constant polynomial with coefficients in $\mathbb{C}$ has a root in $\mathbb{C}$. If we replace the field $\mathbb{C}$ by any other field F, the above discussion applies verbatim, except that the characteristic polynomial of T does not need to have any roots in F.

As a corollary of the above discussion we get the following observation (which is not needed for the solution of the problem). Suppose that n is odd and M is an $n \times n$ matrix with real entries. Then the characteristic polynomial $\det(\lambda I - M)$ is a polynomial of odd degree n with real coefficients. As any odd degree polynomial over the real numbers has a real root, there is a real number a such that $aI - M$ is not invertible. Thus if n is odd, no two matrices can be strongly independent.

Solution to part a): We will think of the matrices M, N as linear transformations $\mathbb{C}^n \longrightarrow \mathbb{C}^n$. From the discussion above, there is a complex number λ such that $Mv = \lambda v$ for some non-zero vector $v \in \mathbb{C}^n$. Let V be the set of all such vectors v (including the zero vector). It is easy to check that V is a subspace of $\mathbb{C}^n$: if $v, w \in V$ then $v + w$ and cv are in V for any complex number c. Now consider any $w \in V$, so $Mw = \lambda w$. Thus $NMw = \lambda Nw$. Since $MN = NM$, we have $M(Nw) = \lambda Nw$, i.e. $Nw \in V$. In other words, the restriction of N to V is a linear transformation from V to itself. Thus there is a complex number τ and a non-zero vector $w \in V$ such that $Nw = \tau w$. It follows that for any numbers a, b, c we have

$$(aI + bM + cN)w = (a + b\lambda + c\tau)w.$$

In particular, if we could find real numbers a, b, c, not all 0, such that $a + b\lambda + c\tau = 0$ then the matrix $aI + bM + cN$ would not be invertible, which would prove that M, N are not strongly independent. Write $\lambda = s + t\mathrm{i}$ and $\tau = p + q\mathrm{i}$. If $t = 0$ then $a = -s$, $b = 1$, $c = 0$ work. If $t \neq 0$, then $a = tp - qs$, $b = q$, $c = -t$ work. This completes our proof that M, N are not strongly independent.

Solution to part b): Consider the matrices

$$M = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad N = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}.$$

We have $M^2 = -I = N^2$ and $MN = -NM$. It follows that $(aI + bM + cN)(aI - bM - cN) = (a^2 + b^2 + c^2)I$. Thus, if not all of a, b, c are 0, then $a^2 + b^2 + c^2 \neq 0$ and $aI + bM + cN$ is invertible with inverse

$$\frac{a}{a^2 + b^2 + c^2} I - \frac{b}{a^2 + b^2 + c^2} M - \frac{c}{a^2 + b^2 + c^2} N.$$

In other words, M and N are strongly independent 4×4 matrices.

Finally, if M, N are strongly independent $n \times n$ matrices and M_1, N_1 are strongly independent $m \times m$ matrices, then the block-diagonal $(n + m) \times (n + m)$ matrices

$$\begin{bmatrix} M & 0 \\ 0 & M_1 \end{bmatrix}, \quad \begin{bmatrix} N & 0 \\ 0 & N_1 \end{bmatrix}.$$

are also strongly independent. It is now straightforward to deduce part b) from this discussion.

Exercise 1. Show directly that the matrix $aI + bM + cN$ in our solution to part b) has determinant $(a^2 + b^2 + c^2)^2$.

Exercise 2. Show that no two 2×2 matrices are strongly independent.

Remark 1. The solution to part b) may appear ad hoc, but there is in fact a deeper underlying reason for it. In order to explain it, we need to recall the concept of the quaternion number system $\mathbb{H}$. $\mathbb{H}$ is a ring (i.e. it has addition and multiplication obeying the usual rules except that multiplication is not commutative) and it contains three elements $\mathrm{i}, \mathrm{j}, \mathrm{k}$ with the following properties:

1. the real numbers $\mathbb{R}$ form a subring of $\mathbb{H}$ and $ax = xa$ for any $a \in \mathbb{R}$ and any $x \in \mathbb{H}$.
2. $\mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = -1$, $\mathrm{ij} = \mathrm{k} = -\mathrm{ji}$
3. Every element of $\mathbb{H}$ can be expressed in a unique way as $a + b\mathrm{i} + c\mathrm{j} + d\mathrm{k}$ for some real numbers a, b, c, d.

Using the above properties we can easily see, for example, that

$$\mathrm{jk} = \mathrm{j}(\mathrm{ij}) = \mathrm{j}(-\mathrm{ji}) = (\mathrm{j}(-\mathrm{j}))\mathrm{i} = -(\mathrm{jj})\mathrm{i} = (-(-1))\mathrm{i} = \mathrm{i}.$$

It turns out that every non-zero element of $\mathbb{H}$ is invertible. In fact, one can easily verify the following identity:

$$(a + b\mathrm{i} + c\mathrm{j} + d\mathrm{k})(a - b\mathrm{i} - c\mathrm{j} - d\mathrm{k}) = a^2 + b^2 + c^2 + d^2.$$

While it is not completely obvious that such a number system $\mathbb{H}$ exists, some basic linear algebra easily leads to a nice construction of $\mathbb{H}$ as a subring of 4×4 matrices over $\mathbb{R}$. In fact, property 3 above means that $\mathbb{H}$ is a 4 dimensional vector space over $\mathbb{R}$ with basis $1, \mathrm{i}, \mathrm{j}, \mathrm{k}$. For any $x \in \mathbb{H}$, left multiplication by x defines a linear transformation $L_x : \mathbb{H} \longrightarrow \mathbb{H}$. Let M_x be the matrix of L_x in the basis $1, \mathrm{i}, \mathrm{j}, \mathrm{k}$. It is straightforward to see that the assignment $x \mapsto M_x$ defines injective ring homomorphisms $\mathbb{H} \longrightarrow \mathrm{M}_4(\mathbb{R})$ from $\mathbb{H}$ into the ring of 4×4 matrices $\mathrm{M}_4(\mathbb{R})$. More explicitly, we have $L_{\mathrm{i}}(1) = \mathrm{i}$, $L_{\mathrm{i}}(\mathrm{i}) = -1$, $L_{\mathrm{i}}(\mathrm{j}) = \mathrm{k}$, $L_{\mathrm{i}}(\mathrm{k}) = -\mathrm{j}$, so the matrix M_{i} is exactly the matrix M from our solution to part b). Similarly, $M_{\mathrm{j}} = N$ and

$$M_{\mathrm{k}} = MN = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} = L.$$

So far we showed that if $\mathbb{H}$ exists then it can be identified with the set of all matrices of the form $aI + bM + cN + dL$, where a, b, c, d are real numbers. Conversely, it is now an easy exercise to see that taking as $\mathbb{H}$ the set of all matrices of the form $aI + bM + cN + dL$, where a, b, c, d are real numbers and setting $\mathrm{i} = M, \mathrm{j} = N, \mathrm{k} = L$, we get a ring and the properties 1-3 above are satisfied. In particular, if at least one of a, b, c, d is not 0 then the matrix

$$aI + bM + cN + dL = \begin{bmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{bmatrix}$$

is invertible.

Exercise 3. Show that the matrix $aI + bM + cN + dL$ above has determinant $(a^2 + b^2 + c^2 + d^2)^2$. Hint: what is the product of this matrix with its transpose?

Remark 2. The above discussion suggest the following generalization of the definition of strong independence. We say that $n \times n$ matrices $M_1, \ldots, M_k$ with real entries are strongly independent if the matrix $aI + a_1M_1 + \ldots + a_kM_k$ is invertible for every choice of real numbers $a, a_1, \ldots, a_k$ which are not all 0. From the discussion at the beginning of our solution, a single matrix M is strongly independent if and only if M does not have real eigenvalues. In particular, when n is odd then every real $n \times n$ matrix has a real eigenvalue (since odd degree polynomials must have a real root), so no collection of

$n \times n$ matrices can be strongly independent. Let us define $\tau(n)$ to be the largest k such that there exist k strongly independent $n \times n$ matrices. Thus $\tau(n) = 0$ if n is odd. It is easy to see that $\tau(2) = 1$ (see Exercise 2 above; can you give an example of a 2×2 matrix with no real eigenvalue?). From Remark 1 we see that $\tau(4) \geq 3$ (the matrices M, N, K are strongly independent). On the other hand, it is a simple generalization of Exercise 2 to show that $\tau(n) < n$ for every n. Thus $\tau(4) = 3$. What about $\tau(6)$? It turns out that $\tau(6) = 1$, but I do not know any simple argument to prove this. More generally, $\tau(2n) = 1$ for every odd number n. I do not know any purely algebraic proof of this fact. The value of $\tau(n)$ is actually known for all n:

Theorem. *Let $n = 2^k m$, where m is odd, and let $k = 4a + b$, where $b \in \{0, 1, 2, 3\}$ (so a, b, m are uniquely determined by n). Then $\tau(n) = 8a + 2^b - 1$.*

For n as in the Theorem, let us write $\mathrm{hr}(n) = 8a + 2^b - 1$. The fact that $\tau(n) \geq \mathrm{hr}(n)$ follows from some purely algebraic results obtained over a 100 years ago by A. Hurwitz (in 1923) and J. Radon (in 1922). The opposite inequality $\tau(n) \leq \mathrm{hr}(n)$ is a much deeper and harder result, proved by J. F. Adams in 1962 using then newly developed tools in algebraic topology (K-theory).

Below we first discuss some ideas connecting strong independence to some questions in topology and then, in a sequence of exercises, we outline an argument showing that $\tau(n) \geq \mathrm{hr}(n)$.

Suppose that $M_1, \ldots, M_k$ are strongly independent $n \times n$ matrices. Then, for any non-zero $v \in \mathbb{R}^n$ the vectors $v, M_1 v, \ldots, M_k v$ are linearly independent. Suppose that v is a unit vector, which we think of as a point on the $(n-1)$-sphere $S^{n-1} = \{(u_1, \ldots, u_n) : u_1^2 + \ldots + u_n^2 = 1\}$. Applying the Gramm-Schmidt orthonormalization process to $v, M_1 v, \ldots, M_k v$ we get vectors $v, T_1(v), \ldots, T_k(v)$ which are orthogonal to each other. The vectors $T_i(v)$ depend continuously on v and therefore each T_i is a unit vector field on S^{n-1}. Thus we constructed k linearly independent (actually orthogonal to each other) unit vector fields on the sphere S^{n-1}. Using deep results in K-theory, J. F. Adams proved in 1962 that the number of linearly independent non-vanishing vector fields on the sphere S^{n-1} can not exceed $\mathrm{hr}(n)$. Hence $\tau(n) \leq \mathrm{hr}(n)$. In particular, applying Adams result when n is odd, we see that on an even dimensional sphere there are no non-vanishing continuous vector fields. This special case is in fact a much older result often called the Hairy Ball Theorem (or Hedgehog Theorem).

Now we outline an argument showing that $\tau(n) \geq \mathrm{hr}(n)$. Recall that an $n \times n$ matrix M with real entries is called orthogonal if $MM^t = I$, where M^t denotes the transpose matrix. Equivalently, M is orthogonal if its columns (rows) are orthogonal to each other and all have length 1 (as vectors in $\mathbb{R}^n$). For example, the matrices M, N, L in Remark 1 are orthogonal. The following exercise suggests an idea to extend the result of Remark 1.

Exercise 4. Suppose that $M_1, \ldots, M_k$ are $n \times n$ orthogonal matrices such that $M_i^t = -M_i$ and $M_i M_j = -M_j M_i$ for any $i \neq j$. Prove that $M_1, \ldots, M_k$ are strongly independent (use the hint in Exercise 3).

Any set of orthogonal matrices which has the properties described in Exercise 4 will be called a Hurwitz-Radon set, or H-R set for short. Thus the set $\{M, N, K\}$ from Remark 1 is an H-R set of 4×4 matrices. If we construct an H-R set of $n \times n$ matrices with $\mathrm{hr}(n)$ elements, the inequality $\tau(n) \geq \mathrm{hr}(n)$ will follow.

Our main tool is an operation on matrices called Kronecker product. If A is an $m \times n$ matrix and B is a $p \times q$ matrix then the Kronecker product $A \otimes B$ is the $mp \times nq$ matrix obtained by replacing every entry $a_{i,j}$ in A with the matrix $a_{i,j} B$. For example,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

and

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Exercise 5. a) Prove that $(A \otimes B)^t = A^t \otimes B^t$.

b) If A_1, A_2 are $n \times n$ matrices and B_1, B_2 are $p \times p$ matrices then $(A_1 \otimes B_1)(A_2 \otimes B_2) = A_1A_2 \otimes B_1B_2$

c) $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.

d) $I_m \otimes I_n = I_{mn}$, where I_k is the $k \times k$ identity matrix.

e) $A \otimes kB = kA \otimes B = k(A \otimes B)$ for any real number k.

f) If A, B are orthogonal then so is $A \otimes B$.

Exercise 6. Let $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, $S = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, and $T = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

a) Show that $A^2 = -I_2$, $S^2 = I_2 = T^2$, $AS = -T = -SA$, $AT = P = -TA$, $ST = A = -TS$.

b) Show that the matrices $I_2 \otimes A \otimes I_2$, $I_2 \otimes S \otimes A$, $T \otimes T \otimes A$, $S \otimes T \otimes A$, $A \otimes S \otimes T$, $A \otimes S \otimes S$, $A \otimes T \otimes I_2$ form an H-R set. Conclude that $\tau(8) = 7$.

Exercise 7. a) Let $\{N_1, \dots, N_7\}$ be an H-R set of seven 8×8 matrices (it exists by part b) of problem 6). Let $\{M_1, \dots, M_k\}$ be an H-R set of $n \times n$ matrices. Let A, S, T be as in problem 6. Prove that the set $\{S \otimes I_n \otimes N_1, \dots, S \otimes I_n \otimes N_7\} \cup \{T \otimes M_1 \otimes I_8, \dots, T \otimes M_k \otimes I_8\} \cup \{A \otimes I_n \otimes I_8\}$ is an H-R set of $k + 8$ matrices of size $16n \times 16n$.

b) Use part a) and induction to show that for every n there exists an H-R set consisting of $\mathrm{hr}(n)$ matrices of size $n \times n$. Conclude that $\tau(n) \geq \mathrm{hr}(n)$ for all n.

Remark. Hurwitz proved that there exists an H-R set of $n-1$ $n \times n$ matrices if and only if $n = 2, 4, 8$. Radon gave an algebraic proof that there is no H-R set of $n \times n$ matrices with more that $\mathrm{hr}(n)$ elements.