Archive of past papers, handouts, and quizzes for

MATH 220, Linear Algebra

Fall 2011, LJB,

version: 10 January 2012.

Source file: arch220fall11.tex

MATH 220 LINEAR ALGEBRA, Fall 2011

Course specification

Laurence Barker, Mathematics Department, Bilkent University,
version: 27 September 2011.

Course Aims: To learn some introductory theory and techniques of linear algebra, and to develop some practical skills of mathematical reasoning.

Course Instructor: Laurence Barker, Office SAZ 129.

Course Assistant: Merve Demirel

Course Texts:

Primary: B. Kolman, D. R. Hill, *Elementary Linear Algebra with Applications*, 9th Edition, (Pearson, 2008).

Secondary: Howard Anton, *Elementary Linear Algebra*, 6th Edition, (Wiley, 1991).

Classes: Tuesdays 14:40 - 15:30 SAZ 01, Thursdays 15:40 - 17:30, SAZ 01.

It is in the nature of any mathematics course that, sometimes, it is impossible to understand everything during the class. To fully grasp the ideas, you must study them regularly on your own, firstly by working through lecture-notes and textbooks, secondly by tackling exercises. It is virtually impossible to pick the ideas up during the two days before an exam. If the exam is only two days away, and if you do not know the material yet, then you should give up. For that reason, there will be no special office hours during the few days before an exam.

Office Hours: Wednesdays 13:40 - 14:30, in my office, Science Faculty Building, A Block, room SA 129.

The Office Hours is important, because it is your main opportunity to have a sustained one-to-one or several-to-one dialogue with me. And it is my opportunity to get some feedback about the course. It is often during office hours that I learn about major difficulties that have been affecting many of the students.

During Office Hours, you may ask me about the homework questions. Office Hours is also an appropriate time to ask me anything else about mathematics, on or off the syllabus.

Class announcements: You will be held responsible for being aware of any announcements made in class, whether or not you were in attendance. That includes announcements about locations of Midterm Exams and any announcements about changes of exam times.

Grading method: Curve.

Exams: The exams are closed-book. Make-ups will be harder than Midterms, and will be granted only if a medical note from a doctor is produced.

- Quizzes, 15%.
- Midterm I, 25%, 3 November, 18:00 - 20:00.
- Midterm II, 25%, 8 December, 18:00 - 20:00.
- Final, 35%.

A minimum of 75 percent attendance is obligatory. Attendance will be measured by counting returned scripts for pop-up quizzes. Failure to hand in scripts for at least 75 percent of the quizzes will result in a grade reduction (B- to C+, or B to B-).

Syllabus: Week number: Monday, subtopic (Primary textbook section number).

- 1: Sept 26, Systems of linear equations, matrices, 1.1 - 1.5.
- 2: Oct 3, Echelon form, nonsingular matrices, 2.1 - 2.3.
- 3: Oct 10, Elementary matrices, LU factorization, 2.3 - 2.5.
- 4: Oct 17, Determinants and applications, 3.1 - 3.5.
- 5: Oct 24, Vector spaces, subspaces, 4.1 - 4.5.
- 6: Oct 31, Linear independence, basis, dimension 4.5 - 4.6.
- 7: Nov 7, Holiday
- 8: Nov 14,Coordinates, homogenous systems, rank, 4.7 - 4.9.
- 9: Nov 21,Inner product spaces, 5.1, 5.2.
- 10: Nov 28, Gram–Schmidt Process, orthogonal complements, 5.3, 5.4.
- 11: Dec 5, Linear transformations 6.1, 6.2.
- 12: Dec 12, Linear transformations, similarity of matrices, 6.3 - 6.5.
- 13: Dec 19, Eigenvalues, eigenvectors, 7.1, 7.2.
- 14: Dec 26, Diagonalization, 7.3.
- 15: Jan 2, Applications of eigenvalues and eigenvectors, 8.1 - 8.3.

Handout 2 for MATH 220

Notes on Determinants and Inverses of Matrices

Fall 2011, Laurence Barker, Mathematics Department, Bilkent University.

Warning: These notes are intended as a reference for an introductory course on linear algebra or an introductory course on group theory. Their purpose is to supply some proofs of some important results: the multiplicative property of determinants, the vanishing of the determinant for singular matrices, and the role of determinants in a formula for the inverse of a square matrix. Illustrative numerical examples can be found in textbooks.

For the purposes of our discussion, it can be understood that, for all the matrices under consideration, the entries are complex numbers.

Let us begin with some preliminary comments on matrix multiplication. Recall that, given positive integers r, s, t and an $r \times s$ matrix A and an $s \times t$ matrix B, then the product AB is the $r \times t$ matrix such that, writing $a_{i,j}$ and $b_{j,k}$ and $c_{i,k}$, respectively, for the (i,j) entry of A and the (j,k) entry of B and the (i,k) entry of AB, we have

$$c_{i,k} = \sum_j a_{i,j}\, b_{j,k} \ .$$

Matrix multiplication is associative. We mean to say, given another positive integer u and a $t \times u$ matrix C, then $(AB)C = A(BC)$. So we can write ABC unambiguously.

We define the **transpose** of A, denoted A^T, to be the $s \times t$ matrix A^T such that the (j,i) entry of A^T is equal to the (i,j)-entry of A. It is easy to see that the transpose of a product is the product of the transposes,

$$(AB)^T = B^T\, A^T \ .$$

Now let A be a square matrix, we mean to say, an $n \times n$ matrix, where n is a positive integer. We say that A is **invertible** or **non-singular** if there exists an $n \times n$ matrix A^{-1} such that $A^{-1}\, A = I = A\, A^{-1}$, where I denotes the identity $n \times n$ matrix. In that case, we call A^{-1} the **inverse** of A. The inverse, if it exists, is unique. Indeed, given $n \times n$ matrices B and C such that $BA = I = AC$ then, using the associative property of matrix multiplication, we have $B = BI = BAC = IC = C$. A slightly weaker characterization of the inverse will appear in Corollary 12, below. When no inverse exists, we say that A is **non-invertible** or **singular**. The following two remarks are obvious.

Remark 1: (The inverse of the transpose is the transpose of the inverse.) *Given an invertible $n \times n$ matrix A, then the $n \times n$ matrix A^T is invertible, and $(A^T)^{-1} = (A^{-1})^T$.*

Remark 2: (The inverse of a product is the product of the inverses.) *Given invertible $n \times n$ matrices A and B, then AB is invertible and $(AB)^{-1} = B^{-1} A^{-1}$.*

Given a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, we define the **determinant** of A to be $\det(A) = ad - bc$. If $\det(A) \neq 0$, then A is invertible. Indeed, by direct calculation again, it is easy to check that

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \ .$$

Conversely, by the next exercise, if A is invertible then $\det(A^{-1})\det(A) = \det(A^{-1}A) = \det(I) = 1$, hence $\det(A) \neq 0$.

Exercise A: By direct calculation, show that

$$\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} e & f \\ g & h \end{bmatrix}\right) = \det\begin{bmatrix} a & b \\ c & d \end{bmatrix} . \det\begin{bmatrix} e & f \\ g & h \end{bmatrix} .$$

Let us write out the definition of the determinant of a 2×2 matrix in a different way. As a briefer notation, we sometimes write the determinant of a 2×2 matrix A as $|A| = \det(A)$. The defining formula for the determinant is

$$\begin{vmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{vmatrix} = a_{1,1}a_{2,2} - a_{1,2}a_{2,1} .$$

Now let A be a 3×3 matrix. Write $a_{i,j}$ for the (i, j) entry of A. We define the **determinant** of A, denoted $\det(A)$ or $|A|$, to be

$$\begin{vmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{vmatrix} = a_{1,1}\begin{vmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{vmatrix} - a_{1,2}\begin{vmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{vmatrix} + a_{1,3}\begin{vmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{vmatrix}$$

$$= a_{1,1}\, a_{2,2}\, a_{3,3} - a_{1,1}\, a_{2,3}\, a_{3,2} - a_{1,2}\, a_{2,1}\, a_{3,3} + a_{1,2}\, a_{2,3}\, a_{3,1} + a_{1,3}\, a_{2,1}\, a_{3,2} - a_{1,3}\, a_{2,2}\, a_{3,1} .$$

Before defining the determinant of an $n \times n$ matrix for an arbitrary positive integer n, we need to introduce the notion of a permutation.

Consider a set X. We define a **permutation** of X to be a bijection from X to X, we mean to say, an invertible function from X to X. Given permutations ρ and σ of X, we write $\rho\sigma$ to denote the composite of ρ and σ. Thus, $\rho\sigma$ is the permutation of X such that $(\rho\sigma)(x) = \rho(\sigma(x))$ for $x \in X$. We write $\mathrm{Sym}(X)$ to denote the set of permutations on X. Usually, we call $\rho\sigma$ the **product** of ρ and σ. We think of $\mathrm{Sym}(X)$ as a set equipped with an operation, called **multiplication**, which sends a pair of elements ρ and σ of $\mathrm{Sym}(X)$ to the element $\rho\sigma$ of X.

Our concern will be with the case where X is replaced by the set $\mathbb{Z}_n^+ = \{1, 2, ..., n-1, n\}$ of positive integers less than or equal to n. We write $S_n = \mathrm{Sym}(\mathbb{Z}_n^+)$. Note that $|S_n| = n!$. Let us introduce a convenient notation for representing elements of S_n. Given mutually distinct elements i_1, i_2, ..., i_{r-1}, i_r of $\mathbb{Z}_n^+$, we write $(i_1, i_2, ..., i_r)$ to denote the element of S_n such that, given $k \in \mathbb{Z}_n^r$, then

$$(i_1, ..., i_n)(k) = \begin{cases} i_{t+1} & \text{if } k = i_t \text{ for some } 1 \leq t < r, \\ i_1 & \text{if } k = i_r, \\ k & \text{otherwise.} \end{cases}$$

The permutation $(i_1, ..., i_r)$ is called an **r-cycle**. As an example, putting $n = 2$, we have

$$S_2 = \{1, (1, 2)\}$$

where 1 denotes the identity function on the set $\mathbb{Z}_2^\times = \{1, 2\}$. Putting $n = 3$, we have

$$S_3 = \{1, (1, 2), (1, 3), (2, 3), (1, 2, 3), (1, 3, 2)\} .$$

The 2-cycles in S_n, called the **transpositions** in S_n, play an especially important role in the theory. These are the elements $(i, j) = (j, i)$ where i and j are distinct elements of $\mathbb{Z}_n^+$. We have $(i,j)(i) = j$ and $(i,j)(j) = (i)$ and $(i,j)(k) = k$ for all the other elements k of $\mathbb{Z}_n^+$. Note that, given a transposition τ in S_n, then $\tau^2 = 1$, in other words, $\tau^{-1} = \tau$.

Two integers are said to have the **same parity** provided they are both even or both odd. They are said to have **opposite parity** provided one of them is even and one of them is odd.

Lemma 3: (Well-definedness of the signature of a permutation.) *For all $n \geq 2$, any element σ of S_n is a product of transpositions. Writing $\sigma = \tau_r ... \tau_1 = \tau'_{r'} ... \tau'_1$ as products of transpositions τ_i and τ'_i, then the integers r and r' have the same parity.*

Proof: For the first part, we argue by induction on n. The case $n = 2$ is clear. Now suppose that $n \geq 3$ and assume that the assertion holds for S_{n-1}. Let $\sigma \in S_{n-1}$. If $\sigma(n) = n$, then we can regard σ as an element of S_{n-1}, hence σ is a product of transpositions. On the other hand, if $\sigma(n) \neq n$ then, introducing the transposition $\tau = (\sigma(n), n)$. and letting $\rho = \tau\sigma$, we have $\rho(n) = n$, hence ρ is a product of transpositions. But $\sigma = \tau\rho$, so σ is a product of transpositions. The first part is established.

Let $\Pi(\sigma) = \{\{u,v\} \subseteq \mathbb{Z}_n^+ : u < v,\, \sigma(u) > \sigma(v)\}$. Let τ be a transposition in S_n, and write $\tau = (i,j)$ with $i < j$. Consider the sets $\{u,v\}$ that belong to exactly one of the sets $\Pi(\sigma)$ or $\Pi(\tau\sigma)$. These are precisely the sets such that

$$\{\{\sigma(u), \sigma(v)\} \in \{(\{i,j\},\, \{i,k\},\, \{k,j\} : i < k < j\} \;.$$

The number of such sets is $2j - 2i - 1$, which is odd. Therefore $|\Pi(\sigma)|$ and $|\Pi(\tau\sigma)|$ have opposite parity. But $|\Pi(1)| = 0$. An inductive argument now shows that if σ is a product of r transpositions, then $|\Pi(\sigma)|$ and r have the same parity. The rider follows. $\square$

For any positive integer n and any element $\sigma \in S_n$, we define

$$\mathrm{sgn}(\sigma) = (-1)^r = \begin{cases} 1 & \text{if } r \text{ is even,} \\ -1 & \text{if } r \text{ is odd.} \end{cases}$$

In the trivial case $n = 1$, the only permutation is the identity function 1, and we understand that $\mathrm{sgn}(1) = 1$. We call $\mathrm{sgn}(\sigma)$ the **signature** of σ. Note that, given elements $\rho, \sigma \in S_n$, then

$$\mathrm{sgn}(\rho\sigma) = \mathrm{sgn}(\rho)\mathrm{sgn}(\sigma) \;.$$

Now let A be an $n \times n$ matrix with (i,j)-entry $a_{i,j}$ for $i,j \in \mathbb{Z}_n^+$. We define the **determinant** of A, denoted $\det(A)$ or $|A|$, to be

$$\det(A) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma)\, a_{\sigma(1),1} ... a_{\sigma_n, n} \;.$$

Proposition 4: *With the notation above, suppose that two rows of A are the same, or that two columns of A are the same. That is to say, for some i and j with $i \neq j$, we have $a_{i,k} = a_{j,k}$ for all k, or we have $a_{k,i} = a_{k,j}$ for all k. Then* $\det(A) = 0$.

Proof: Suppose that row i and row j are the same, with $i \neq j$. Consider the transposition $\tau = (i,j)$. We can arrange the elements of S_n in pairs, where elements σ and σ' of S_n are partners provided $\sigma' = \tau\sigma$, or equivalently, $\sigma = \tau\sigma'$. When σ and σ' are partners, we have $\mathrm{sgn}(\sigma) + \mathrm{sgn}(\sigma') = 0$ and

$$a_{\sigma(1),1} ... a_{\sigma(n),n} = a_{\sigma'(1),1} ... a_{\sigma'(n),n} \;.$$

So $\det(A) = 0$. The case where two columns are the same can be dealt with similarly, by pairing σ with $\sigma\tau$, or alternatively, it can be deduced by considering the transpose of A. $\square$

Theorem 5: (Multiplicative property of determinants.) *Let n be a positive integer and let A and B be $n \times n$ matrices. Then* $\det(AB) = \det(A)\det(B)$.

Proof: Let $C = AB$ and let $a_{i,j}$ and $b_{j,k}$ and $c_{i,k}$ denote, respectively, the (i,j) entry of A, the (j,k) entry of B, the (i,k) entry of C. Since $c_{i,k} = \sum_j a_{i,j} b_{j,k}$, we have

$$\det(C) = \sum_{\pi} \operatorname{sgn}(\pi)\, c_{\pi(1),1} ... c_{\pi(n),n} = \sum_{\pi} \operatorname{sgn}(\pi) \big(\sum_{j_1,...,j_n} a_{\pi(1),j_1}\, b_{j_1,1} ... a_{\pi(n),j_n}\, b_{j_n,n} \big)$$

summed over all the elements $\pi \in S_n$ and $j_1, ..., j_n \in \mathbb{Z}_n^+$. Defining

$$\gamma(j_1, ..., j_n) = \sum_{\pi} \operatorname{sgn}(\pi) a_{\pi(1),j_1}\, b_{j_1,1} ... a_{\pi(n),j_n}\, b_{j_n,n}$$

and changing the order of the summation, we have

$$\det(C) = \sum_{j_1,...,j_n} \gamma(j_1, ..., j_n) \; .$$

On the other hand,

$$\det(A)\det(B) = \big(\sum_{\sigma} \operatorname{sgn}(\sigma) a_{\sigma(1),1} ... a_{\sigma(n),n} \big) \big(\sum_{\rho} \operatorname{sgn}(\rho) b_{\rho(1),1} ... b_{\rho(n),n} \big)$$

$$= \sum_{\sigma,\rho} \operatorname{sgn}(\rho\sigma) a_{\sigma(1),1} ... a_{\sigma(n),n}\, b_{\rho(1),1} ... b_{\rho(n),n}$$

summed over all the elements $\sigma, \rho \in S_n$. Changing the order of the multiplication,

$$b_{\rho(1),1} ... b_{\rho(n),n} = b_{\rho\sigma(1),\sigma(1)} ... b_{\rho\sigma(n),\sigma(n)} \; .$$

For each σ, the product $\pi = \rho\sigma$ runs over the elements of S_n as ρ runs over the elements of S_n. Therefore

$$\det(A)\det(B) = \sum_{\sigma,\pi} \operatorname{sgn}(\pi) a_{\pi(1),\sigma(1)}\, b_{\sigma(1),1)} ... a_{\pi(n),\sigma(n)}\, b_{\sigma(n),n)} = \sum_{\sigma} \gamma(\sigma(1), ..., \sigma(n)) \; .$$

It suffices to show that $\gamma(j_1, ..., j_n) = 0$ when the integers j_1, ..., j_n are not mutually distinct. Suppose that $j_u = j_v$ with $u \neq v$. Of course, the assumption implies that $n \geq 2$. Consider the transposition $\tau = (u, v)$. Much as in the previous argument, we can arrange the elements of S_n in pairs, partnering elements π and π' of S_n when $\pi' = \tau\pi$, whence $\operatorname{sgn}(\pi) + \operatorname{sgn}(\pi') = 0$ and $a_{\pi(1),j_1}\, b_{j_1,1} ... a_{\pi(n),j_n}\, b_{j_n,n} = a_{\pi'(1),j_1}\, b_{j_1,1} ... a_{\pi'(n),j_n}\, b_{j_n,n}$. We deduce that $\gamma(j_1, ..., j_n) = 0$, as required. $\square$

Corollary 6: (The determinant of the inverse is the inverse of the determinant.) *Given an invertible $n \times n$ matrix A, then* $\det(A) \neq 0$ *and* $\det(A^{-1}) = (\det(A))^{-1}$.

Proof: We have $\det(A^{-1})\det(A) = \det(A^{-1}A) = \det(I) = 1$. $\square$

Exercise B, for the mathematically inclined: The quaternions are an extension of the complex numbers. They have the form $q = t + ix + jz + ky$ where t, x, y, z are real numbers which uniquely determine q. We define a multiplication operation on the quaternions such that $i^2 = j^2 = k^2 = ijk = -1$ and $qr = rq$ for all real numbers r. Let $\mathbb{H}$ denote the set of quaternions, let $\mathrm{Mat}_2(\mathbb{C})$ denote the set of 2×2 matrices and let ρ be the function $\mathbb{H} \to \mathrm{Mat}_2(\mathbb{C})$ such that

$$\rho(q) = t \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + x \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} + y \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} + z \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} .$$

(a) Show that $\rho(qq') = \rho(q)\rho(q')$. Deduce that multiplication of quaternions is associative.

(b) Let $N(q) = t^2 + x^2 + y^2 + z^2$. Show that $N(q) = \det(\rho(q))$ and $N(qq') = N(q)N(q')$.

(c) A natural number n is said to be a **sum of four squares** provided $n = t^2+x^2+y^2+z^2$ for some integers t, x, y, z. Using part (b), show that, if natural numbers n and m are sums of four squares, then nm is a sum of four squares. (This conclusion is due to Euler. Subsequently, in 1771, Lagrange made use of this to prove that every natural number is a sum of four squares.)

We define the **permutation matrix** associated with an element $\sigma \in S_n$ to be the $n \times n$ matrix $A(\sigma)$ whose (i, j) entry is 1 when $i = \sigma(j)$ and whose (i, j) entry is zero otherwise. The next two remarks are obvious.

Remark 7: *Given an elements $\rho, \sigma \in S_n$, then $A(\rho)A(\sigma) = A(\rho\sigma)$.*

Remark 8: *Given an element $\sigma \in S_n$, then $\det(A(\sigma)) = \mathrm{sgn}(\sigma)$.*

We shall also need the following technical lemma.

Lemma 9: *Let $\sigma \in \S_n$ and $i, j \in \mathbb{Z}_n^+$ such that $i = \sigma(j)$. Let M be the matrix obtained from $A(\sigma)$ by deleting the i-th row and the j-th column. Then $(-1)^{i+j}\det(M) = \mathrm{sgn}(\sigma)$.*

Proof: As permutations, let $\alpha = (n, n-1, ..., i+1, i)$ and $\beta = (j, j+1, ..., n-1, n)$ and $\pi = \alpha\sigma\beta$. We have $\pi(n) = n$, and the matrix M is obtained from $A(\pi)$ by deleting the n-th row and the n-th column, so we can regard π as an element of S_n with permutation matrix M. Using the latest two remarks and the multiplicative property of determinants

$$\det(M) = \det(A(\pi)) = \det(A(\alpha)A(\beta)A(\gamma))$$

$$= \det(A(\alpha)) \det(A(\sigma) \det(A(\beta)) = \mathrm{sgn}(\alpha) \,\mathrm{sgn}(\sigma) \,\mathrm{sgn}(\beta) \;.$$

But $\alpha = (n, n-1)(n-1, n-2)...(i+1, i)$ as a product of $n-i$ transpositions, so $\det(\alpha) = n-i$. Similarly, $\det(\beta) = n - j$. The required conclusion follows. □

The next result characterizes determinants in a recursive way that is sometimes useful for practical calculation. Let A be an $n \times n$ matrix. Again, we write $a_{i,j}$ for the (i, j) entry of A. Of course, in the case $n = 1$, we have $\det(A) = a_{1,1}$. For $n \geq 2$, we can express $\det(A)$ in terms of the determinants of some $(n-1) \times (n-1)$ matrices. Let $M_{i,j}$ be the $(n-1) \times (n-1)$ matrix obtained from A by deleting the i-th row and the j-th column. Let $A_{i,j} = (-1)^{i+j}\det(M_{i,j})$.

Theorem 10: *With the notation above, let C be the $n \times n$ matrix with (i, j) entry $A_{i,j}$. Then $AC = \det(A)I = CA$. In other words, for all $k \in \mathbb{Z}_n^+$, we have*

$$a_{k,1}\, A_{1,k} + a_{k,2}\, A_{2,k} + ... + a_{k,n}\, A_{n,k} = A_{k,1}\, a_{1,k} + A_{k,2}\, a_{2,k} ... + A_{k,n}\, a_{n,k} = \det(A)$$

and for all $i, j \in \mathbb{Z}_n^+$ with $i \neq j$, we have

$$a_{i,1}\, A_{1,j} + a_{i,2}\, A_{2,j} + ... + a_{i,n}\, A_{n,j} = A_{i,1}\, a_{1,j} + A_{i,2}\, a_{2,j} ... + A_{i,n}\, a_{n,j} = 0 \;.$$

Proof: We have

$$a_{k,1}\,A_{1,k} + a_{k,2}\,A_{2,k} + ... + a_{k,n}\,A_{n,k} = (-1)^{k+1} a_{k,1} \det(M_{k,1}) + ... + (-1)^{k+n} a_{k,n} \det(M_{k,n})$$
$$= \sum_{\sigma} s_k(\sigma)\, a_{1,\sigma(1)} ... a_{n,\sigma(n)}$$

where $s_k(\sigma) = \pm 1$ and $s_k(\sigma)$ depends on σ and possibly on k, but not on A. To show that $s_k(\sigma) = \mathrm{sgn}(\sigma)$, we may assume that $A = A(\sigma)$. Writing $k = \sigma(j)$ then, by the latest lemma,

$$s_k(\sigma) = (-1)^{k+j} \det(M_{k,j}) = \mathrm{sgn}(\sigma) \; .$$

We have have now established that

$$a_{k,1}\,A_{1,k} + a_{k,2}\,A_{2,k} + ... + a_{k,n}\,A_{n,k} = \det(A) \; .$$

The other asserted equality for $\det(A)$ holds by a similar argument or, alternatively, it can be deduced by considering the transpose of A.

For all i and j in $\mathbb{Z}_n^+$, we have

$$a_{i,1}\,A_{1,j} + a_{i,2}\,A_{2,j} + ... + a_{i,n}\,A_{n,j} = (-1)^{j+1} a_{i,1} \det(M_{j,1}) + ... + (-1)^{j+n} a_{i,n} \det(M_{j,n})$$

Supposing now that $i \neq j$ then, since each of the matrices appearing in the right-hand expression has been obtained by deleting row j from A, the value of the right-hard expression will not change if we replace row j of A with row i of A. But then, by the previous paragraph, the right-hand expression is the determinant of a matrix whose i-th row and j-th row are the same. Hence, via Proposition 4, $a_{i,1}\,A_{1,j} + ... + a_{i,n}\,A_{n,j} = 0$. The remaining asserted equality can be demonstrated by a similar argument or, alternatively, by considering transposes. □

Corollary 11: *Given a square matrix A, then A is invertible if and only if* $\det(A) = 0$. *In that case, the (i, j) entry of A^{-1} is, in the notation above,* $A_{i,j}/\det(A)$.

Proof: This follows immediately from Corollary 6 and the latest theorem. □

Corollary 12: *Let A and B be $n \times n$ matrices. Suppose that $AB = I$ or $BA = I$. Then A and B are invertible and $A^{-1} = B$.*

Proof: The hypothesis, combined with the multiplicative property of determinants, implies that $\det(A)\det(B) = \det(I) = 1$. Hence $\det(A) \neq 0$ and A is invertible. The uniqueness property of the inverse now implies that $A^{-1} = B$. □

Recall that the three elementary row operations are: multiplying a row by a non-zero scalar factor, interchanging two rows, adding one row to another row.

Exercise C: Find a method for calculating the determinant of a square matrix based on using elementary row operations to convert the matrix to upper triangular form. (Hint: the determinant of an upper triangular matrix is easy to calculate. Consider the determinants of the matrices representing the three kinds of row operation. Alternative hint: the method can be found in textbooks.)

Comment for mathematics students: All of the above material holds for matrices over an arbitrary field. The set S_n, equipped with the multiplication we imposed, is a group called the **symmetric group of degree** n. The function $\mathrm{sgn} : S_n \to \{\pm 1\}$ is a group homomorphism. The kernel $A_n = \{\sigma \in S_n : \mathrm{sgn}(\sigma) = 1\}$ is called the **alternating group of degree** n**.** The groups S_n and A_n crop up in many different contexts of application.

Handout 3 for MATH 220

Notes on Diagonalization and Change of Basis

Laurence Barker, Mathematics Department, Bilkent University,
version: 30 October 2011.

Warning: These notes are intended as a reference for an introductory course on linear algebra. Their purpose is to summarize the rationale behind the method of diagonalization. Illustrative numerical examples can be found in textbooks.

Toy Problem: Let x_n and y_n be the number of female tribbles and male tribbles, respectively, on day n. Every night, each female tribble gives birth to 2 tribbles, both of them male, and each male tribble gives birth to 2 tribbles, both of them female. Tribbles never die. It is given that $x_0 = 3$ and $y_0 = 5$. Give a formula for the number of tribbles on day n. (Tribbles are small cute alien creatures which look like fluffy tennis-balls and which reproduce very fast. See the episode *The Trouble with Tribbles* of the original *Star Trek* television series.)

Answer: We have $x_n = 4.3^n - (-1)^n$ and $y_n = 4.3^n + (-1)^n$.

Proof 1: We argue by induction on n. The case $n = 0$ is trivial. Now

$$x_{n+1} = x_n + 2y_n \;, \qquad y_{n+1} = 2x_n + y_n$$

for all natural numbers n. Assume, inductively, that the assertion holds for x_n and y_n. Then

$$x_{n+1} = (4.3^n - (-1)^n) + 2(4.3^n + (-1)^n) = 4.3^{n+1} - (-1)^{n+1}$$

and similarly for y_{n+1}, as required. $\square$

Proof 2: Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ and $f_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $f_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Then $Af_1 = 3f_1$ and $Af_2 = -f_2$, hence

$$\begin{bmatrix} x_n \\ y_n \end{bmatrix} = A \begin{bmatrix} x_{n-1} \\ y_{n-1} \end{bmatrix} = A^n \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = A^n \begin{bmatrix} 3 \\ 5 \end{bmatrix} = A^n(4f_1 + f_2) = 4A^n f_1 + A^n f_2$$

$$= 4.3^n \begin{bmatrix} 1 \\ 1 \end{bmatrix} + (-1)^n \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4.3^n - (-1)^n \\ 4.3^n + (-1)^n \end{bmatrix} \; . \quad \square$$

The second proof contains the seeds of a systematic method, which we shall explain below.

Change of basis:

Let us begin by recalling how linear maps can be represented by matrices. Let V be finite-dimensional vector space over a field F and let $\alpha : V \to V$ be a linear map. Writing $n = \dim(V)$, let us choose a basis $\mathcal{E} = \{e_1, ..., e_n\}$ for V. The matrix A representing α with respect to $\mathcal{E}$ is the $n \times n$ matrix A such that, given vectors $x = x_1 e_1 + ... + x_n e_n$ and $y = y_1 e_1 + ... + y_n e_n$ with $\alpha(x) = y$, then $A\underline{x} = \underline{y}$ where $\underline{x} = (x_1, ..., x_n)$ and $\underline{y} = (y_1, ..., y_n)$. In other words, letting $a_{i,j}$ be the (i, j) entry of A, then $y_i = \sum_j a_{i,j}\, x_j$.

It is important to note that the matrix A depends not only on α but also on the choice of basis $\mathcal{E}$. In many contexts of application, it is helpful to change the basis so that α is represented by a different matrix which is easier to work with.

Let $\mathcal{F} = \{f_1, ..., f_n\}$ be another basis for V. Since $\mathcal{E}$ and $\mathcal{F}$ are bases, there exist unique scalars $t_{i,j}$ and $s_{j,i}$ such that

$$f_j = \sum_j t_{i,j}\, e_i \ , \qquad\qquad e_i = \sum_j s_{j,i}\, f_j$$

Writing $x = x_1 e_1 + ... + x_n e_n = \sum_i x_i\, e_i$ and $x = \sum_j x'_j\, f_j$, then $x = \sum_{i,j} t_{i,j} x'_j\, e_i$. By the uniqueness of coordinates with respect to a given basis,

$$x_i = \sum_j t_{i,j}\, x'_j \ , \qquad\qquad x'_j = \sum_i s_{j,i}\, x_i \ .$$

The coordinate vectors $\underline{x} = (x_1, ..., x_n)$ and $\underline{x'} = (x'_1, ..., x'_n)$ represent x with respect to $\mathcal{E}$ and $\mathcal{F}$, respectively. We have

$$\underline{x} = T\underline{x'}$$

where T is the matrix with (i, j)-entry $t_{i,j}$. Similarly, $\underline{x'} = S\underline{x}$, where S is the matrix with (j, i)-entry $s_{j,i}$. Plainly, $ST = I = TS$ where I denotes the identity $n \times n$ matrix. In other words, $S = T^{-1}$ and

$$\underline{x'} = T^{-1}\underline{x} \ .$$

We call T the **coordinate transformation matrix** from $\mathcal{F}$-coordinates to $\mathcal{E}$-coordinates. Evidently, T^{-1} is the coordinate transformation matrix in the other direction, $\mathcal{E}$-coordinates to $\mathcal{F}$-coordinates.

Still letting A be the matrix representing α with respect to $\mathcal{E}$, now let B be the matrix representing α with respect to $\mathcal{F}$. The equation $y = \alpha(x)$ can be expressed in coordinate form as $\underline{y} = A\underline{x}$ and as $\underline{y'} = B\underline{x'}$. But $\underline{y'} = T^{-1}\underline{y} = T^{-1}A\underline{x} = T^{-1}AT\underline{x'}$. It follows that

$$B = T^{-1}AT \ , \qquad\qquad A = TBT^{-1} \ .$$

These observations motivate the following definition. Given $n \times n$ matrices A and B, then A is said to be **similar** to B provided there exists an invertible $n \times n$ matrix T such that $A = TBT^{-1}$. The following remark is easy to check.

Remark: *Let A, B, C be $n \times n$ matrices. Then A is similar to A. If A is similar to B, then B is similar to A. If A is similar to B and if B is similar to C, then A is similar to C.*

In other words, similarity of $n \times n$ matrices is an equivalence relation.

We can now clear up a loose-end from earlier on in the course. We define the **determinant** of the linear map α to be the scalar

$$\det(\alpha) = \det(A)$$

where A is a matrix representing α with respect to some basis. The determinant of α is well-defined, independently of the choice of basis, thanks to the following remark.

Remark: *Given similar $n \times n$ matrices A and B, then* $\det(A) = \det(B)$.

Proof: Write $A = TBT^{-1}$. Then $\det(A) = \det(T)\det(B)\det(T)^{-1} = \det(B)$. □

Exercise: Given an $n \times n$ matrix A, writing $a_{i,j}$ for the (i, j)-entry of A, we define the **trace** of A to be $\mathrm{tr}(A) = \sum_{i=1}^{n} a_{i,i}$. Let B be an $n \times n$ matrix similar to A. Show that $\mathrm{tr}(A) = \mathrm{tr}(B)$.

In view of the latest exercise, we can define the **trace** of α to be

$$\operatorname{tr}(\alpha) = \operatorname{tr}(A)$$

where A is a matrix representing α. Indeed, the exercise implies that $\operatorname{tr}(\alpha)$ is well-defined.

Diagonal representation of linear maps

An $n \times n$ matrix A is said to be **diagonal** provided the (i,j)-entry is zero whenever $i \neq j$. In that case, letting $a_{i,i}$ denote the (i,i)-entry, we write $A = \operatorname{diag}(a_{1,1}, ..., a_{n,n})$. Diagonal matrices tend to be very easy to work with. For instance, if A is diagonal, then $A^m = \operatorname{diag}(a_{1,1}^m, ..., a_{n,n}^m)$ for any positive integer m, moreover, A is invertible if and only if each $a_{i,i} \neq 0$ and, in that case, $A^{-1} = \operatorname{diag}(a_{1,1}^{-1}, ..., a_{n,n}^{-1})$.

We say that A is **diagonalizable** provided A is similar to a diagonal matrix. In other words, A is diagonalizable provided there exists a diagonal matrix B and an invertible matrix T such that $A = TBT^{-1}$. In that case, one way of calculating A^m is to make use of the equality $A^m = TB^mT^{-1}$. Also, if A is invertible, then B is invertible, and $A^{-1} = TB^{-1}T^{-1}$.

Again, let V be an n-dimensional vector space over a field F, and let $\alpha : V \to V$ be a linear map. We say that α is **diagonal** provided α is represented by a diagonal matrix with respect to some basis. Below, we shall describe a method for finding a diagonal matrix representing a given diagonal linear map.

First, we need a definition. Given a non-zero vector $f \in V$ and a scalar $\lambda \in F$ such that $\alpha(f) = \lambda f$, we call f an **eigenvector** of α with **eigenvalue** λ.

Remark: *A scalar $\lambda \in F$ is an eigenvalue of α if and only if $\det(\alpha - \lambda I) = 0$, where I denotes the identity map on V.*

Proof: Both of the specified conditions are plainly equivalent to the condition that the equation $(\alpha - \lambda I)x = 0$ has a non-zero solution $x \in V$. $\square$

The equation $\det(\alpha - \lambda I) = 0$ is called the **characteristic** equation of the linear map α.

Choosing a basis $\mathcal{E}$ of V and letting A be the matrix representing α with respect to $\mathcal{E}$, the characteristic equation of α can be rewritten as $\det(A - \lambda I) = 0$, where I now denotes the identity matrix. Sometimes, we call this equation the **characteristic equation** of the matrix A. In an evident way, we can also speak of the **eigenvectors** and **eigenvalues** of A.

At last, we can explain the idea behind the second proof pertaining to the toy problem above. If we can find a basis $\mathcal{F} = \{f_1, ..., f_n\}$ of V such that each f_j is an eigenvector for α, say $\alpha(f_j) = \lambda_j f_j$, then the matrix B representing α with respect to $\mathcal{F}$ is the diagonal matrix $B = \operatorname{diag}(\lambda_1, ..., \lambda_n)$, where λ_i is the eigenvalue associated with the eigenvector f_j.

In applications, two kinds of scenario often arise. In one of them, a diagonal linear map α is given, the matrix A representing α with respect to some basis $\mathcal{E}$ has been determined, and the task is to find another basis $\mathcal{F}$ such that α is represented by a diagonal matrix B with respect to $\mathcal{F}$. Letting T be the coordinate transformation matrix from $\mathcal{F}$-coordinates to $\mathcal{E}$-coordinates, then $A = TBT^{-1}$. In the other kind of scenario, a diagonalizable matrix A is given, and we seek a diagonal matrix B and an invertible matrix T such that $A = TBT^{-1}$. This is really the same problem as before, and we can understand α to be the linear map on F^n such that α is represented by A with respect to the standard basis of F^n.

The procedure is as follows:

Step 1: Find the eigenvalues λ_1, ..., λ_n, which are the solutions to the polynomial equation $\det(A - \lambda I) = 0$ (possibly with repeated solutions). Then $B = \operatorname{diag}(\lambda_1, ..., \lambda_n)$.

Step 2: Find the corresponding eigenvectors f_i by solving the equation $(A - \lambda I)f_j = 0$ (taking care to find m linearly independent eigenvectors associated with an eigenvalue that has multiplicity m as a repeated solution). The matrix T is the matrix whose j-th column is the coordinate vector representing f_j with respect to $\mathcal{E}$.

Return of the toy problem

As a first little example, let us deal with the above toy problem systematically, using the method that we have described. The eigenvalues of the matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ are the solutions to the equation

$$0 = \det(A - \lambda I) = \det \begin{bmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{bmatrix} = (1-\lambda)^2 - 2^2 = \lambda^2 - 2\lambda - 3 \ .$$

The solutions are $\lambda_1 = 3$ and $\lambda_2 = -1$. Write $f_1 = (u, v)$ as a coordinate vector with respect to the standard basis $\mathcal{E} = \{(1,0),(0,1)\}$ of $\mathbb{R}^2$. To find f_1, we solve

$$0 = \begin{bmatrix} 1-\lambda_1 & 2 \\ 2 & 1-\lambda_1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

which yields $u = v$. So we can put $f_1 = (1,1)$. A similar calculation with λ_2 in place of λ_1 yields a solution $f_2 = (-1,1)$ as a coordinate vector with respect to $\mathcal{E}$. Taking the columns of T to be the $\mathcal{E}$-coordinates of the eigenvectors, $T = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, whence $T^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$. Then $A = TBT^{-1}$ and

$$A^n = TB^nT^{-1} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}^n \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3^n & 3^n \\ -(-1)^n & (-1)^n \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 3^n + (-1)^n & 3^n - (-1)^n \\ 3^n - (-1)^n & 3^n + (-1)^n \end{bmatrix} \ .$$

As a check, we note that, putting $n = 1$, the latest equality reduces to the definition of A. Finally, we recover the answer

$$\begin{bmatrix} x_n \\ y_n \end{bmatrix} = A^n \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 3^n + (-1)^n & 3^n - (-1)^n \\ 3^n - (-1)^n & 3^n + (-1)^n \end{bmatrix}\begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 4.3^n - (-1)^n \\ 4.3^n + (-1)^n \end{bmatrix} \ .$$

Actually, it was not really necessary to calculate T^{-1}. We could, instead, have argued more along the lines that we presented earlier.

Exercise: Let A be an $n \times n$ matrix with eigenvalues λ_1, ..., λ_n, up to multiplicity. Thus $\det(A - \lambda I) = (\lambda_1 - \lambda)...(\lambda_n - \lambda)$. Show that $\mathrm{tr}(A) = \lambda_1 + ... + \lambda_n$.

A more difficult problem

The following problem is somewhat similar to the one discussed above, but it is of interest because repeated eigenvalues appear.

Problem: A machine has three possible states, labelled 1, 2, 3. For distinct states i and j, if the machine is in state i at time $t = n$, then the probablity of the machine being in state j

at time $t = n + 1$ is 1/4. Suppose that the machine is in state 1 at time $t = 0$. What is the probability of the machine being in state 1 at time $t = n$?

Solution: Consider the matrix $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$. Let $p_i(n)$ denote the probability of the machine being in state i at time $t = n$. Then, writing column vectors as rows for convenience, $4(p_{n+1}(1), p_{n+1}(2), p_{n+1}(3)) = A(p_n(1), p_n(2), p_n(3))$ and $(p_0(1), p_0(2), p_0(3)) = (1, 0, 0)$.

The matrix $A - I$ has three identical non-zero rows and hence has nullity 2. So 1 appears twice as an eigenvalue of A. The matrix $A - 4I$ is non-invertible because the sum of its rows is zero, hence 4 is an eigenvalue of A. Therefore, the eigenvalues of A are $\lambda_1 = 1$ and $\lambda_2 = 1$ and $\lambda_3 = 4$. We can put $f_1 = (1, -1, 0)$ and $f_2 = (1, 0, -1)$ because these two vectors are eigenvectors with associated eigenvalue 1 and the set $\{f_1, f_2\}$ is linearly independent. We can put $f_3 = (1, 1, 1)$ as an eigenvector with associated eigenvalue 4. Thus $Af_1 = f_1$ and $Af_2 = f_2$ and $Af_3 = 4f_3$. We have $(p_0(1), p_0(2), p_0(3)) = (f_1 + f_2 + f_3)/3$, hence

$$\begin{aligned}(p_n(1), p_n(2), p_n(3)) &= (A/4)^n(f_1 + f_2 + f_3)/3 = (f_1 + f_2 + 4^n f_3)/3.4^n \\ &= (4^n + 2,\ 4^n - 1,\ 4^n - 1)/3.4^n \ .\end{aligned}$$

In conclusion, $p_n(1) = (4^n + 2)/3.4^n = (1 + 1/4^{n-1})/3$.

Let us mention that, as a variant of the proof that the eigenvalues of A are 1, 1, 4, we could have observed that the matrices $A - I$ and $A - 4I$ are non-invertible, hence the eigenvalues are λ, 1, 4 for some scalar λ. Then, using an exercise above, we could have noted that $\lambda + 1 + 4 = \mathrm{tr}(A) = 2 + 2 + 2 = 6$, hence $\lambda = 1$. As another alternative, more routine (rather boring, in fact), we could have made the calculation

$$\det(A - \lambda I) = \begin{vmatrix} 2-\lambda & 1 & 1 \\ 1 & 2-\lambda & 1 \\ 1 & 1 & 2-\lambda \end{vmatrix} = (2-\lambda)\begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} - \begin{vmatrix} 1 & 1 \\ 1 & 2-\lambda \end{vmatrix} + \begin{vmatrix} 1 & 2-\lambda \\ 1 & 1 \end{vmatrix}$$

$$= (2-\lambda)(\lambda^2 - 4\lambda + 3) - (1-\lambda) + (-1+\lambda) = -\lambda^3 + 6\lambda^2 - 9\lambda + 4 = (1-\lambda)(1-\lambda)(4-\lambda) \ .$$

When does the diagonalization procedure work?

Again, we let α be a linear map $V \to V$, where V is an n-dimensional vector space over a field F. We choose a basis $\mathcal{E} = \{e_1, ..., e_n\}$, and we let A be the matrix representing α with respect to $\mathcal{E}$. The next remark is obvious.

Remark: *The following three conditions are equivalent:*
(a) *The linear map α is diagonal.*
(b) *The matrix A is diagonalizable.*
(c) *There exists a bases $\mathcal{F} = \{f_1, ..., f_n\}$ of V such that each f_i is an eigenvector of α.*

The next result gives a sufficient criterion for those three equivalent conditions to hold.

Proposition: *Suppose that α has n mutually distinct eigenvalues λ_1, ..., λ_n in F. Let f_1, ..., f_n be corresponding eigenvectors, in order. Then $\{f_1, ..., f_n\}$ is a basis for F^n. In particular, α is diagonal and A is diagonalizable.*

Proof: For a contradiction, suppose that $\{f_1, ..., f_n\}$ is not linearly independent. Write

$$\mu_1 f_1 + ... + \mu_n f_n = 0$$

where each $\mu_i \in F$, some $\mu_i \neq 0$, and the positive integer $m = |\{i : \mu_i \neq 0\}|$ is as small as possible. Renumbering the f_i if necessary, we may assume that

$$\mu_1 f_1 + ... + \mu_m f_m = 0$$

and that $\mu_i \neq 0$ for all $1 \leq i \leq m$. Plainly, $m \geq 2$. But

$$\lambda_1 \mu_1 f_1 + ... + \lambda_m \mu_m f_m = \alpha(\mu_1 f_1 + ... + \mu_m f_m) = \alpha(0) = 0 \ .$$

Multiplying the first equation by λ_m and then subtracting it from the second equation, we obtain

$$(\lambda_1 - \lambda_m)\mu_1 f_1 + ... + (\lambda_{m-1} - \lambda_m)\mu_{m-1} f_{m-1} = 0 \ .$$

But $m-1 \geq 1$ and all the coefficients $(\lambda_i - \lambda_m)\mu_i$ are non-zero. This contradicts the minimality of m. $\square$

It is not hard to see that, if A is diagonalizable, then the above procedure for expressing A in the form $A = TBT^{-1}$ can always be applied successfully.

Sometimes, A may not be diagonalizable over F, yet A may be diagonalizable as a matrix over a larger field. An interesting example of this is the matrix

$$R_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

which, of course, represents an anticlockwise rotation of the Euclidian plane $\mathbb{R}^2$ through an angle of θ. Regarding R_θ as a matrix over $\mathbb{R}$, then plainly R_θ is not diagonalizable unless θ is an integer multiple of π (in which case, $R_\theta = \pm I$.) But let us now regard R_θ as a matrix over the field of complex numbers $\mathbb{C}$. Thus, R_θ now represents a linear map $\mathbb{C}^2 \to \mathbb{C}^2$. The characteristic equation of R_θ is

$$0 = \det(R_\theta - \lambda I) = \begin{vmatrix} c - \lambda & -s \\ s & c - \lambda \end{vmatrix} = (c - \lambda)^2 + s^2 = \lambda^2 - 2c\lambda + 1$$

where $c = \cos(\theta)$ and $s = \sin(\theta)$. The solutions are $\lambda_1 = c + is = e^{i\theta}$ and $\lambda_2 = c - is = e^{-i\theta}$. It is easy to check that the corresponding eigenvectors are $f_1 = (1, -i)$ and $f_2 = (1, i)$. So $T = \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}$. We have $\det(T) = 2i$, hence $T^{-1} = \dfrac{1}{\det(T)} \begin{bmatrix} i & -1 \\ i & 1 \end{bmatrix} = \dfrac{1}{2} \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix}$. Therefore

$$R_\theta = T \operatorname{diag}(\lambda_1, \lambda_2)\, T^{-1} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix} \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} \ .$$

One advantage of working over $\mathbb{C}$ rather than $\mathbb{R}$ is that, for any $n \times n$ matrix A over $\mathbb{C}$, there always exist scalars $\lambda_i \in \mathbb{C}$ such that

$$0 = \det(A - \lambda I) = (\lambda_1 - \lambda)...(\lambda_n - \lambda) \ .$$

Indeed, the Fundamental Theorem of Algebra asserts that, given complex numbers a_{n-1}, ..., a_0, then there exist complex numbers λ_1, ..., λ_n such that, for all complex numbers λ, we have

$$\lambda^n + a_{n-1}\lambda^{n-1} + ... + a_1\lambda + a_0 = (\lambda - \lambda_1)...(\lambda - \lambda_n) \ .$$

It is perhaps rather surprising that, even over $\mathbb{C}$, non-diagonalizable matrices exist.

Proposition: *Let A be a 2×2 matrix over $\mathbb{C}$. Then A is non-diagonalizable if and only if A is similar to the matrix* $\begin{bmatrix} a & b \\ 0 & a \end{bmatrix}$ *where $a, b \in \mathbb{C}$ with $b \neq 0$.*

Proof: Let B be the specified matrix. The characteristic equation of B is $0 = \det(C - \lambda I) - (a - \lambda)^2$. So a is the unique eigenvalue of C. It is now easy to see that the eigenvectors of C are precisely the vectors having the form $(x, 0)$ where x is a non-zero complex number. These vectors do not span the vector space $\mathbb{C}^2$, so C is not diagonalizable, and any matrix similar to C is non-diagonalizable.

Conversely, suppose that A is non-diagonalizable. By the Fundamental Theorem of Algebra, A has an eigenvalue $a \in \mathbb{C}$. On the other hand, by the previous Proposition, A cannot have two distinct eigenvalues. That is to say, a must be the unique eigenvalue of A. Choose an eigenvector f_1 of A, and choose any vector f_2 in $\mathbb{C}^2$ such that $\{f_1, f_2\}$ is a basis for $\mathbb{C}^2$. Letting α be the linear map $\mathbb{C}^2 \to \mathbb{C}^2$ represented by A with respect to the standard basis of $\mathbb{C}^2$, then $\alpha(f_1) = Af_1 = af_1$ and $\alpha(f_2) = bf_1 + df_2$ for some $b, d in \mathbb{C}$. So, letting B be the matrix representing α with respect to the basis $\{f_1, f_2\}$, we have $B = \begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$. But B has a unique eigenvalue and the characteristic equation of B is $0 = \det(B - \lambda I) = (a - \lambda)(d - \lambda)$, hence $a = d$. Furthermore, B cannot be a diagonal matrix, so $b \neq 0$. □

For many kinds of matrix that appear frequently in contexts of application — symmetric matrices or unitary matrices, for instance — there are results which guarantee diagonalizability. But the proof of the next theorem illustrates a scenario where failure of diagonalizibility arises naturally. Actually, the fastest way to prove the theorem is by induction, nevertheless, the argument we present is an entertaining exercise in the theory developed above.

Incidentally, the following proof is also an illustration of the use of theory as opposed to calculation. We shall be arguing simply by making deductions from conceptual principles, without carrying out any substantial manipulations of written symbols.

Theorem: *Let a, b, c be complex numbers with $a \neq 0$. Let x_0, x_1, ... be an infinite sequence of complex numbers such that $ax_{n+2} + bx_{n+1} + cx_n = 0$ for all natural numbers n. If the quadratic equation $a\lambda^2 + b\lambda + c = 0$ has two distinct solutions λ_1 and λ_2, then there exist complex numbers u_1 and u_2 such that $x_n = u_1\lambda_1^n + u_2\lambda_2^n$ for all n. If the quadratic equation has a unique non-zero solution λ, then there exist complex numbers μ and ν such that $x_n = (\mu + n\nu)\lambda^n$ for all n. If 0 is the unique solution to the quadratic equation, then $x_n = 0$ for all $n \geq 2$.*

Proof: We may assume that $a = 1$, since otherwise we can replace b and c with b/a and c/a, respectively. We may also assume that $c \neq 0$, since otherwise the required conclusion is trivial. We can now understand x_n to be defined for all integers n, with x_{-1}, x_{-2}, ... and x_2, x_3, ... recursively determined by x_0 and x_1 via the equality $x_{n+1} + bx_n + cx_{n-1} = 0$. Let us rewrite the equality as $\begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix} = A \begin{bmatrix} x_n \\ x_{n-1} \end{bmatrix} = A^n \begin{bmatrix} x_1 \\ x_0 \end{bmatrix}$ where $A = \begin{bmatrix} -b & -c \\ 1 & 0 \end{bmatrix}$ and n is any integer. This makes sense because $\det(A) = c \neq 0$ and A^n is defined for all n.

Let V be the vector space over $\mathbb{C}$ consisting of the functions $\mathbb{Z} \to \mathbb{C}$, with the evident addition and scalar multiplication operations. We shall be making use of the observation that the function $n \mapsto x_n$ can be regarded as a vector in V.

The eigenvalues of A are the complex numbers λ such that $\det(A - \lambda I) = 0$, in other words, $\lambda^2 + b\lambda + c = 0$. It is easy to see that, given any eigenvalue λ of A, then the corresponding eigenvectors are precisely the vectors $(\lambda y, y)$ where y is a non-zero complex number.

Suppose that the equation has two distinct solutions λ_1 and λ_2. Note that λ_1 and λ_2 are both non-zero because $\lambda_1\lambda_2 = c \neq 0$. The matrix A is diagonalizable by a proposition above. (Alternatively, we can argue that A must be diagonalizable because the eigenvectors $(\lambda_1, 1)$ and $(\lambda_2, 1)$ comprise a basis for $\mathbb{C}^2$.) Therefore $A = T\,\mathrm{diag}(\lambda_1, \lambda_2)\,T^{-1}$ for some invertible matrix T, and $A^n = T\,\mathrm{diag}(\lambda_1^n, \lambda_2^n)\,T^{-1}$ for all integers n. Observing that T is independent of n, it is not hard to see that, as a vector in V, the function $n \mapsto x_n$ is a linear combination of the functions $n \mapsto \lambda_1^n$ and $n \mapsto \lambda_2^n$.

It remains to deal with the case where the quadratic equation has a unique solution λ. We have $\lambda \neq 0$ because $\lambda^2 = c \neq 0$. All the eigenvectors of A are the scalar multiples of the vector $(\lambda, 1)$. These vectors do not span $\mathbb{C}^2$, so A cannot be diagonalizable. By another proposition above, A is similar to the matrix $B = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$. A straightforward inductive argument yields $B^n = \begin{bmatrix} \lambda^n & n\lambda^{n-1} \\ 0 & \lambda^n \end{bmatrix}$ for all integers n. Writing $A = TBT^{-1}$, then $A^n = TB^nT^{-1}$ and

$$\begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix} = T \begin{bmatrix} \lambda_0^n & n\lambda_0^{n-1} \\ 0 & \lambda_0^n \end{bmatrix} T^{-1} \begin{bmatrix} x_1 \\ x_0 \end{bmatrix} .$$

But the functions $n \mapsto \lambda^{n-1}$ and $n \mapsto \lambda^n$ and $n \mapsto \lambda^{n+1}$ are all scalar multiples of each other, and similarly for the functions $n \mapsto n\lambda^{n-1}$ and $n \mapsto n\lambda^n$ and $n \mapsto n\lambda^{n+1}$. So the function $n \mapsto x_n$ is a linear combination of the functions $n \mapsto \lambda^n$ and $n \mapsto n\lambda^n$. $\square$

MATH 220: LINEAR ALGEBRA. Midterm 1.
LJB and UM, 3 November 2011, Bilkent University.

Below is a record of the exam questions, together with solutions and comments. The duration of the exam was two hours. All the questions had equal weight.

1: Let A and B be $n \times n$ matrices. Without using the theory of determinants, show that if AB is nonsingular then A and B must be nonsingular.

2: Show that the following matrix is nonsingular. Express it as a product of elementary matrices.

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 5 & 3 \\ -2 & 0 & -3 \end{bmatrix}.$$

3: Consider the following matrix:

$$\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ 1 & t & t^2 \end{bmatrix}.$$

Suppose that $ae - bd \neq 0$. Show that there are at most two values of t such that $\det(A) = 0$.

4: Let

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix}.$$

Find a lower triangular matrix $\mathbf{L}$ and an upper triangular matrix $\mathbf{U}$ such that $\mathbf{A} = \mathbf{LU}$. Using the LU factorization, solve the following equation

$$\mathbf{A} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

5: Let A be $n \times n$ matrix, and let B be a matrix obtained from A by interchanging two rows. Show that $\det(A) = -\det(B)$. Deduce that if two rows of A are the same as each other then $\det(A) = 0$.

Solutions and comments for Midterm 1:

1: We use the following standard theorem: letting U be an $n \times n$ matrix, and supposing there exists an $n \times n$ matrix V such that $UV = I$ or $VU = I$, then $UV = I = VU$ and, furthermore, V is unique. Recall that U is said to be non-singular when such V exists and, in that case, we write $U^{-1} = V$.

Putting $C = B(AB)^{-1}$, then $AC = I$. Hence, via the theorem, A is non-singular. Similarly, B is non-singular.

Alternative: We use the following: letting U be an $n \times n$ matrix, letting μ be the function $\mathbb{R}^n \to \mathbb{R}^n$ or $\mathbb{C}^n \to \mathbb{C}^n$ associated with A and supposing that μ is injective or surjective, then μ is bijective. Recall, U is said to be non-singular when it satisfies the hypothesis.

Let α, β, γ be the functions associated with A, B, AB respectively. Since AB is non-singular, the theorem implies that the composite function $\gamma = \alpha\beta$ is bijective. Therefore α is surjective and β is bijective. Another application of the theorem yields the required conclusion.

Comment: The two theorems above are essentially the same, and the two arguments are the same. It was intended that the candidates would simply appeal to one or the other version of the theorem. The few candidates who succeeded in answering this question did not use the theorem directly, but nevertheless displayed commendable insight by implicitly reproducing a proof of the theorem using elementary matrix operations.

The course and textbook are based on an approach whereby matrix algebra is introduced as a procedural formalism and discussion of the underlying concepts of a vector space and a linear map are postponed. Although that does have some advantages, the candidates' responses indicate a peculiar consequence of the early emphasis on method at the expense of theory.

2: Replacing row r_3 with $2r_1 + r_3$, we have

$$E_1 A = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 5 & 3 \\ 0 & -2 & -1 \end{bmatrix} \qquad \text{where} \qquad E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} .$$

Replacing r_2 with $2r_3 + r_2$, we have

$$E_2 E_1 A = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & 1 \\ 0 & -2 & -1 \end{bmatrix} \qquad \text{where} \qquad E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} .$$

Replacing r_3 with $2r_2 + r_3$, we have

$$E_3 E_2 E_1 A = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{where} \qquad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} .$$

Replacing r_1 with $r_1 + r_2$, we have

$$E_4 ... E_1 A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{where} \qquad E_4 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} .$$

Replacing r_1 with $-2r_3 + r_1$, we have

$$E_5...E_1A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{where} \qquad E_5 = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Replacing r_2 with $-r_3 + r_2$, we have

$$E_6...E_1A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{where} \qquad E_6 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Hence

$$A = (E_6...E_1)^{-1} = E_1^{-1}...E_6^{-1} =$$

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Comment: The factorization is not unique. The method yields many other answers. Many candidates succeeded in reducing A to the identity matrix using elementary row operations, but failed to specify the associated elementary matrices E_j. Several candidates who did specify the associated elementary matrices neglected to pass to their inverses E_j^{-1}.

MATH 220: LINEAR ALGEBRA. Midterm 2.
LJB and UM, 8 December 2011, Bilkent University.

Below is a record of the exam questions, together with solutions and comments. The duration of the exam was two hours. All the questions had equal weight.

1: Show that, if $\{\vec{u}, \vec{v}, \vec{w}\}$ is a basis for a finite dimensional vector space V, then $\{\vec{u} - 2\vec{v} + 3\vec{w},\ 2\vec{u} + \vec{v} - \vec{w},\ \vec{u} - \vec{v} + \vec{w}\}$ is also a basis for V.

2: Let V be the vector space of all functions from $\mathbb{R}$ to $\mathbb{R}$ with usual definitions of addition and scalar multiplication:

$$(f \oplus g)(x) = f(x) + g(x), \qquad (c \odot f)(x) = cf(x), \quad \text{where } c \text{ is a scalar.}$$

Show that,
a) if W_1 is the set of all **even** functions (i.e. $f(x) = f(-x)$ for all $x \in \mathbb{R}$) in V,
b) if W_2 is the set of all **odd** functions (i.e. $f(x) = f(-x)$ for all $x \in \mathbb{R}$) in V,
both W_1 and W_2 are subspace of V.

3: Find the dimension of the real vector space

$$\text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \\ x \end{bmatrix}, \begin{bmatrix} 4 \\ 6 \\ 8 \\ 2x \end{bmatrix} \right\}$$

as a subset of $\mathbb{R}^4$, where x is a real number. (Hint: The answer depends on x.)

4: Let $\{\vec{e}_1, ..., \vec{e}_m\}$ and $\{\vec{f}_1, ..., \vec{f}_n\}$ be linearly independent subsets of a vector space V, and suppose that $\text{span}\{\vec{e}_1, ..., \vec{e}_m\} \cap \text{span}\{\vec{f}_1, ..., \vec{f}_n\} = \{\vec{0}\}$. Show that the set $\{\vec{e}_1, ..., \vec{e}_m, \vec{f}_1, ..., \vec{f}_n\}$ is linearly independent.

5: Let $\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 1 \end{bmatrix}$. Find the rank and nullity of $\mathbf{A}$.

Solutions and comments for Midterm 2:

1: Let $T = \{\vec{u} - 2\vec{v} + 3\vec{w},\ 2\vec{u} + \vec{v} - \vec{w},\ \vec{u} - \vec{v} + \vec{w}\}$. We must show that any vector $\vec{t} \in V$ can be written uniquely as a linear combination of the elements of T. The set $S = \{\vec{u}, \vec{v}, \vec{w}\}$ is a basis for V, so there exist unique real numbers b_1, b_2, b_3 such that

$$\vec{t} = b_1\vec{u} + b_2\vec{v} + b_3\vec{w}\ .$$

But the matrix $\begin{bmatrix} 1 & 2 & 1 \\ -2 & 1 & -1 \\ 3 & -1 & 1 \end{bmatrix}$ is non-singular, because its determinant is

$$1(1-1) - 2(-2+3) + 1(2-3) = -3 \neq 0\ .$$

In other words, the system of equations

$$a_1 + 2a_2 + a_3 = b_1\ , \qquad -2a_1 + a_2 - a_3 = b_2\ , \qquad 3a_1 - a_2 + a_3 = b_3$$

has a unique solution in a_1, a_2, a_3 for any given b_1, b_2, b_3. Therefore, as required, $\vec{t}$ can be written uniquely in the form

$$\begin{aligned} \vec{t} &= (a_1 + 2a_2 + a_3)\vec{u} + (-2a_1 + a_2 - a_3)\vec{v} + (3a_1 - a_2 + a_3)\vec{w} \\ &= a_1(\vec{u} - 2\vec{v} + 3\vec{w}) + a_2(2\vec{u} + \vec{v} - \vec{w}) + a_3(\vec{u} - \vec{v} + \vec{w})\ . \end{aligned}$$

Alternatively, one can show separately that T spans V and that T is linearly independent.

Or as another variation, one can observe that, since $|T| = 3 = \dim(V)$, the spanning property of T is equivalent to the linear independence property of T. Hence, it suffices to show only one of those two properties.

2: Let f and g be vectors in W_2, and let c be a scalar. To show that W_2 is a subspace of V, we must check that $f \oplus g$ and $c \odot f$ belong to W_2. We have

$$(f \oplus g)(-x) = f(-x) + g(-x) = -f(x) - g(x) = -(f(x) \oplus g(x)) = -(f \oplus g)(x),$$

$$(c \odot f)(-x) = cf(-x) = -cf(x) = -(c \odot f)(x).$$

Hence $f \oplus g \in W_2$ and $c \odot f \in W_2$. This completes the proof that W_2 is a subspace of V. The proof of the conclusion for W_1 is similar.

3: By routine methods: We shall show that the dimension of the span is 3 when $x = 5$ and it is 2 otherwise. The dimension is the rank of the matrix

$$\begin{bmatrix} 1 & 1 & 2 & 4 \\ 1 & 2 & 3 & 6 \\ 1 & 3 & 4 & 8 \\ 1 & 4 & x & 2x \end{bmatrix}.$$

Elementary row operations do not change the rank. Applying elementary row operations, we can replace the matrix with

$$\begin{bmatrix} 1 & 1 & 2 & 4 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & x-5 & 2x-10 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Evidently the rank is as asserted above.

By direct argument: Let v_1, v_2, v_3, v_4 be the four specified vectors, in order. Since $v_4 = 2v_3$, we have $\mathrm{span}\{v_1, v_2, v_3, v_4\} = \mathrm{span}\{v_1, v_2, v_3\}$. Plainly, the set $\{v_1, v_2\}$ is linearly independent. The equality

$$\lambda_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

has a unique solution, namely $\lambda_1 = \lambda_2 = 1$. So, if $x = 5$, then the the equality $\lambda_1 v_1 + \lambda_2 v_2 = v_3$ has a solution, namely $\lambda_1 = \lambda_2 = 1$, hence $\dim \mathrm{span}\{v_1, ..., v_4\} = 2$. On the other hand, if $x \neq 5$ then the equality has no solution in λ_1 and λ_2, hence $\dim \mathrm{span}\{v_1, ..., v_4\} = 3$.

4: Let a_1, ..., a_m and b_1, ..., b_n be real numbers and suppose that $\sum_i a_i e_i + \sum_j b_j f_j = 0$. We are to show that each $a_i = 0$ and each $b_j = 0$. Now

$$\sum_i a_i e_i = -\sum_j b_j f_j \in \mathrm{span}\{e_1, ..., e_m\} \cap \mathrm{span}\{f_1, ...f_n\} = \{0\} .$$

So $\sum_i a_i e_i = 0$ and $\sum_j b_j f_j = 0$. But $\{e_1, ..., e_n\}$ is linearly independent, so each $a_i = 0$. Similarly, each $b_j = 0$.

Comments: Some common mistakes are listed below.

4.A: Many candidates wrote down suitable equations, but with absent or incorrect indications as to the logical relationships between the equations, for instance,

"$\sum_{i=1}^n a_i e_i = 0$ when $a_1 = a_2 = ... = a_n = 0$".

Some candidates just wrote down loads of equations connected by mysterious arrows. As has been stressed in class, that does not constitute a deductive argument, To convey a mathematical argument clearly and unambiguously, one should use complete, grammatically correct sentences.

4.B: The crux of the argument is to explain why the equality $\sum_i a_i e_i + \sum_j b_j f_j = 0$ implies the equalities $\sum_i a_i e_i = 0$ and $\sum_j b_j f_j = 0$. Very many candidates gave no explanation at all. Many candidates failed to adequately explain how they made use of the hypothesis $\mathrm{span}\{e_1, ...\} \cap \mathrm{span}\{f_1, ...\} = \{0\}$. One candidate wrote along the lines:

"None of the e_i and no combination of the e_i is an f_i or a combination of the f_i."

That does just about succeed in conveying the idea, although it is clumsy and not quite correct: the zero vector is a linear combination of the e_i and it is also a linear combination of the f_j. However, two candidates expressed variants of the assertion:

"None of the e_i is in the span of the f_j and none of the f_j is in the span of the e_i."

That weakening of the hypothesis is insufficient. A counter-example is the case

$$e_1 = (1, 0, 0, 0, 0) , \qquad e_2 = (0, 1, 0, 0, 0) ,$$

$$f_1 = (1, 1, 1, -1, 0) , \quad f_2 = (1, 1, 0, 1, -1) , \quad f_3 = (1, 1, -1, 0, 1) .$$

Here, $\{e_1, e_2\}$ and $\{f_1, f_2, f_3\}$ are linearly independent and the condition in the latest quote is satisfied, but $\{e_1, e_2, f_1, f_2, f_3\}$ is not linearly independent since $3e_1 + 3e_2 - f_1 - f_2 - f_3 = 0$.

4C: A few candidates argued that $\{e_1, ..., e_m, f_1, ..., f_i\}$ and $\{f_{i+1}, ..., f_n\}$ are linearly independent for all i. But, to do that successfully, either one must include the condition

$$\text{span}\{e_1, ..., e_n, f_1, ..., f_i\} \cap \text{span}\{f_{i+1}, ..., f_m\} = \{0\}$$

as part of the inductive assumption, or else one must make use of the condition $\text{span}\{e_1, ...\} \cap \text{span}\{f_1, ...\} = \{0\}$ in some other way. But neither of those two approaches escapes the need to deal with the crux of the problem. Thus, this inductive approach is a *red herring*, and it does not make the problem any easier.

5: Plainly, any three of the four columns are linearly independent as vectors in $\mathbb{R}^4$, so the rank is at least 3. But the sum of the columns is the zero vector, so the rank is exactly 3. Therefore the nullity is $4 - 3 = 0$.

As an alternative solution, it is easy to see that a vector (x_1, x_2, x_3, x_4) belongs to the null space if and only if $x_1 = x_2 = x_3 = x_4$. So the nullity is 1. It follows that the rank is $4 - 1 = 3$.

The question can also be done in a routine way by using elementary row operations to reduce to a matrix in echelon form.

Archive of Final Exam Questions

1: Let $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, and $\mathbf{B} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, and let $\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$ satisfy $\mathbf{PB} = \mathbf{PA}$.
a) Show that $p_{11} = p_{21} = 0$.
b) Use part (a) to prove that $\mathbf{A}$ cannot be similar to $\mathbf{B}$.

2: The following vectors form a basis for the null space of certain 5×6 matrix $\mathbf{A}$:

$$\vec{v}_1 = (-7,\, 3,\, 1,\, 0,\, 0,\, 0), \quad \vec{v}_2 = (5,\, -1,\, 0,\, 1,\, 0,\, 0), \quad \vec{v}_3 = (1,\, -3,\, 0,\, 0,\, -2,\, 1).$$

Determine the reduced row echelon form of $\mathbf{A}$.

3: Let $\mathbf{A}$ and $\mathbf{B}$ be $m \times n$, and $n \times m$ matrices respectively.
a) Let $\mathbf{E}$ be the reduced row echelon form of $\mathbf{A}$. If $\mathbf{AB} = \mathbf{I}_m$, show that $\mathbf{E}$ does not have any rows of zeros.
b) If $m > n$, show that $\mathbf{AB}$ cannot be $\mathbf{I}_m$.

4: Let V be a vector space with bases $\{\vec{e}_1, \vec{e}_2, \vec{e}_3, \}$, and $\{\vec{f}_1, \vec{f}_2, \vec{f}_3, \}$, and $\vec{e}_1 = \vec{f}_2 + \vec{f}_3$, and $\vec{e}_2 = \vec{f}_1 + \vec{f}_3$, and $\vec{e}_3 = \vec{f}_1 + \vec{f}_2$. Let L be the linear map $L : V \to V$ such that $L(\vec{f}_1) = 2\vec{f}_1$, and $L(\vec{f}_2) = 4\vec{f}_2$, and $L(\vec{f}_3) = 6\vec{f}_3$. Find the matrix representing L with respect to the basis $\{\vec{e}_1, \vec{e}_2, \vec{e}_3, \}$.

MATH 220: LINEAR ALGEBRA. Final Retake.
LJB, September 2012, Bilkent University.

The duration of the exam is two hours. All the questions have equal weight.

Questions 2, 3, 4, 5 use notation from the preceding questions.

1: Let $A = \begin{bmatrix} -1 & 1 & 1 \\ -4 & 4 & 1 \\ -4 & 2 & 3 \end{bmatrix}$. Find the inverse A^{-1} by first finding the cofactor matrix.

2: Find the solutions λ_1, λ_2, λ_3 to the equation $\det(A - \lambda I) = 0$, where λ is a real number and I is the identity 3×3 matrix.

3: For each $j \in \{1, 2, 3\}$, find non-zero solutions to the equation $A \begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix} = \lambda_j \begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix}$.

3: Let $P = \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix}$. Calculate the matrix $P^{-1}AP$.

5: Using the fact that $P^{-1}AP$ is a diagonal matrix, find another way of calculating A^{-1}.

6: Let $\{(a_1, a_2, a_3),\ (b_1, b_2, b_3),\ (c_1, c_2, c_3)\}$ be a basis in $\mathbb{R}^3$ such that, applying the Gram-Schmidt Process to obtain an orthonormal basis, the obtained basis is $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$. What can be deduced about the real numbers a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2, c_3?

Quizzes

1: Solve $x + 2y + 3z = 6$, $-7y - 4z = 2$, $y + 2z = 4$.

2: Solve $a + b + c = 0$, $a + b = 3$, $b + c = 1$ using Gaussian elimination.

3: Solve $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$ using LU decomposition.

4: How many elements of S_n are there? How many of them have even signature?

5: Find $\dim(V)$ where $V = \{\lambda_1 s_1 + \lambda_2 s_2 + \lambda_3 s_3 + \lambda_4 s_4 : \lambda_1, \lambda_2, \lambda_3, \lambda_4 \in \mathbb{R}\}$ when

$$s_1 = (1, 1, 2, 2)\ , \qquad s_2 = (1, 1, 3, 5)\ , \qquad s_3 = (0, 0, 1, 3)\ , \qquad s_4 = (2, 2, 6, 10)\ .$$

6: Using elementary row operations, find the dimension of the space $\{x : Ax = 0\}$ where

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 9 & 1 \\ 0 & 1 & 6 & -3 \\ 0 & 1 & 1 & 1 \end{bmatrix}\ .$$

7: Find a basis for the subspace $\{(x, y, z) : x + y + z = 0\}$ of $\mathbb{R}^3$.

8: Let $f_1 = e_1 + e_2 + e_3$, $f_2 = e_2 + e_3$, $f_1 = e_3$. Suppose that

$$xe_1 + ye_2 + ze_3 = x' f_1 + y' f_2 + z' f_3\ .$$

Express (x, y, z) in terms of (x', y', z'). Express (x', y', z') in terms of (x, y, z).

9: Find an invertible matrix P and a diagonal matrix D such that $\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} = PDP^{-1}$.

10: Find the eigenvalues and eigenvectors of the matrix $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.