Targil 11 – once again, linear algebra

1. Let $A_1A_2\ldots A_n$ be a polygon inscribed in circle. Consider a skew symmetric $n\times n$ matrix (a_{ij}), such that for $i < j$, $a_{ij} = A_iA_j$. Prove that the rank of this matrix is not greater than 2.

Solution. Assume that the circle is unit circle with center at 0 (it just divides the matrix by radius and doesn't affect the rank). Then $A_i = (\cos(2\varphi_i), \sin(2\varphi_i))$.
WLOG, the points go clockwise.
Then $a_{ij} = \sin(\varphi_i - \varphi_j) = \sin(\varphi_i)\cos(\varphi_j) - \sin(\varphi_j)\cos(\varphi_i)$.
Both matrixes $\{\sin(\varphi_i)\cos(\varphi_j)\}$ and $\{\sin(\varphi_j)\cos(\varphi_i)\}$ are of rank 1, hence their difference is of rank 2.
Remark. For the case of 4 points, the determinant is (see targil 2 problem 5): $(A_1A_2\cdot A_3A_4 - A_1A_3\cdot A_2A_4 + A_1A_4\cdot A_2A_3)^2$, and that is equal to 0, so for inscribed quadrilateral $A_1A_2\cdot A_3A_4 + A_1A_4\cdot A_2A_3 = A_1A_3\cdot A_2A_4$. This fact is called Ptolemy's theorem, so problem 1 is sometimes called generalized Ptolemy's theorem.

2. Let A, B, C be $n\times n$ square matrices. Prove that

$$\mathrm{rk}(AB) + \mathrm{rk}(BC) \le \mathrm{rk}(ABC) + \mathrm{rk}(B).$$

First solution. In other words,

$$\mathrm{rk}(AB) - \mathrm{rk}(B) \le \mathrm{rk}(ABC) - \mathrm{rk}(BC).$$

Denote $V = \ker B$, $W = \ker BC$. Obviously $V \subset W$.
For every linear transformation rk = n – dim(ker), therefore the statement may be rewritten as follows:

$$\dim V - \dim\ker(AB) \le \dim W - \dim\ker(ABC)$$

Let $v_1, v_2, \ldots, v_k$ be the basis of $\ker A$.
Complete it to the basis $v_1, v_2, \ldots, v_k, v_{k+1}, \ldots, v_{k+l}$ of $\ker(AB)$.
By adding some more vector we can make the basis for $\ker(ABC)$:

$$v_1, v_2, \ldots, v_k, v_{k+1}, \ldots, v_{k+l}, v_{k+l+1}, \ldots, v_{k+l+m}.$$

Here k, l, m are nonnegative integers.
We claim that a non-zero linear combination of $Av_{k+l+1}, Av_{k+l+2}, \ldots, Av_{k+l+m}$ is not in V. Indeed, if $a_1Av_{k+l+1} + \ldots + a_mAv_{k+l+m} = A(a_1v_{k+l+1} + \ldots + a_mv_{k+l+m})$ is in $V = \ker B$,

thus $a_1v_{k+l+1} + \ldots + a_mv_{k+l+m}$ is in ker(AB), so it is a linear combination of $v_1,\ldots, v_{k+l}$, but that is impossible since $v_1, \ldots , v_{k+l+m}$ are linearly independent.
Therefore, if $u_1, u_2, \ldots, u_r$ is a basis of V, then $u_1, u_2, \ldots, u_r, Av_{k+l+1}, \ldots, Av_{k+l+m}$ form a linearly independent system in W. Hence $r + m \leq \dim W$.
We wanted to prove that:

$$\dim V - \dim \ker(AB) \leq \dim W - \dim \ker(ABC)$$

In our new notation, that is

$$r - (k + l) \leq \dim W - (k + l + m)$$
$$r + m \leq \dim W$$

QED.

Second solution. This proof will be much shorter, but it uses some higher mathematics, namely quotient spaces. Recall, that if X is a linear subspace of linear space Y, then Y can be divided into equivalence classes: two vectors are equivalent, if their difference is in X. The set of those equivalence classes forms a linear space, which is called quotient space and denoted Y/X.

Like in the first solution, we shall transform the claim into form:

$$\dim V - \dim \ker(AB) \leq \dim W - \dim \ker(ABC)$$

Where $V = \ker(B)$, $W = \ker(BC)$. Also denote $V' = \ker(AB)$, W' = ker(ABC). Then the claim may be rewritten as follows:

$$\dim W' - \dim V' \leq \dim W - \dim V$$

A maps space W' into space W. If $w\in W'$ and $Aw\in V$ then $w\in V'$.

So, if $w_1, w_2 \in W'$ and $Aw_1 - Aw_2 \in V$, then $w_1 - w_2 \in V'$.

Therefore A induces an injective linear map from W'/V' to W/V. Hence

$$\dim(W'/V') \leq \dim(W/V).$$

The LHS is dim W' – dim V', and the RHS is dim W – dim V.

3.* **(a)** A linear operator A over $\mathbb{C}^n$ can be considered as a linear operator A_r over $\mathbb{R}^{2n}$, because $\mathbb{C}^n$ is a $2n$-dimensional space over $\mathbb{R}$. Prove that $|\det(A)|^2 = \det(A_r)$.

(b) Formulate and prove a more general claim, about finite field extension (field $\mathbb{C}$ is an extension of field $\mathbb{R}$ of degree 2).

Solution. (a) We shall apply Gauss method to simplify the determinant computation of the complex matrix A, and see how will the determinant of the real matrix be transformed in the process.
Permutation of two rows in the complex matrix, that will multiply the complex determinant by -1, will result in permutation of two pairs of rows in the real matrix which won't change its determinant.
Subtracting the multiple of one row from another row in real matrix will result in subtracting linear combinations of two rows from two different rows in the real matrix, so both determinants will be preserved.
Same operations will behave in the same way on columns.
Complex matrix can be diagonalized by those operations. The elements on the diagonal will be x_1+iy_1, $x_2+iy_2,\ldots, x_n+iy_n$. The real matrix, at the same time, will become a block matrix $\begin{pmatrix} x_1 & -y_1 \\ y_1 & x_1 \end{pmatrix}, \begin{pmatrix} x_2 & -y_2 \\ y_2 & x_2 \end{pmatrix}, \ldots, \begin{pmatrix} x_n & -y_n \\ y_n & x_n \end{pmatrix}$.
The determinant of the real matrix is product of determinants of the blocks, hence the statement becomes obvious.

(b) Let K[α] be a separable finite field extension over field K. That means α is an algebraic number over K, its minimal polynomial over K is $p(x)$ which is of degree n, and it has n distinct roots in algebraic closure of K (but not in K).

Questions*: Is it true that irreducible polynomial of degree n over a field always has n distinct roots in its algebraic closure? Is it true that the finite field extension is always generated by one number?

So, in simple words K[α] is a set of polynomials of degree less than n in α. These polynomials have natural sums, differences, products and divisions (except by 0) which follow from the relation $p(\alpha)$.
A polynomial $p(x)$ has n distinct roots in the algebraic closure: $\alpha_1=\alpha$, α_2, α_3, …, α_n.
So each number $q(\alpha)$ in K[α] has n distinct conjugate numbers, itself included:

$$q(\alpha_1),\ q(\alpha_2),\ \ldots\ ,\ q(\alpha_n)$$

Product of these n numbers will be called the norm of $q(\alpha)$.

Example. $\mathbb{C} = \mathbb{R}[i]$, it is field extension of degree 2 over $\mathbb{R}$. Any number in $\mathbb{C}$ can be represented as $a + bi$, polynomial of degree 1 in i.
The minimal polynomial $x^2 + 1$ has two roots, i and $-i$. Each element $a + bi$ has a norm $a^2 + b^2$, which is a product of two conjugate numbers, $a + bi$ and $a - bi$.

Now we can formulate the generalization.

Theorem. Let A be a matrix / linear operator over K[α]^m. It can be considered as A_K a linear operator over K^{mn}, because K[α] is an n-dimensional linear space. Then the norm of det(A) equals det(A_K).

Proof. Like in (a), the theorem is easily reduced to 1-dimensional case by Gauss method, so we won't repeat it. But here, the one dimensional case is nonobvious. Multiplication by α is a linear operator over K[α].
In the basis 1, α, α^2, … α^{n-1} it looks as follows:

$$\begin{pmatrix} 0 & 0 & ... & 0 & a_o \\ 1 & 0 & ... & 0 & a_1 \\ 0 & 1 & ... & 0 & a_2 \\ ... & ... & ... & ... & ... \\ 0 & 0 & ... & 1 & a_{n-1} \end{pmatrix}$$

Here the last column contains minus the coefficients of the minimal polynomial of α, which is $p(x) = x^n - a_{n-1}x^{n-1} - \ldots - a_2x^2 - a_1x - a_0$.
Since it is hard to guess eigenvalues of that matrix, we shall take the transposed matrix which has the same eigenvalues (see targil 2, problem 3b), and use it for the guessing. For any root α_k of the minimal polynomial,

$$\begin{pmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1 \\ a_o & a_1 & a_2 & ... & a_{n-1} \end{pmatrix} \begin{pmatrix} 1 \\ \alpha_k \\ \alpha_k^2 \\ ... \\ \alpha_k^{n-1} \end{pmatrix} = \begin{pmatrix} \alpha_k \\ \alpha_k^2 \\ \alpha_k^3 \\ ... \\ \alpha_k^n \end{pmatrix}$$

So, that is an eigenvector and α_k is an eigenvalue and all eigenvalues are different. So the matrix of α is diagonalizable, and has eigenvalues α_1, α_2, …, α_n. Therefore the matrix of $q(\alpha)$ = q(matrix of α) and its eigenvalues are $q(\alpha_1)$, $q(\alpha_2)$, …, $q(\alpha_n)$. Hence the determinant is the product of those.

4. Consider matrix equation $AX - XB = C$, where A, B, C are given n×n matrixes, and X is an unknown n×n matrix. Show that the solution of the equation exists and unique if and only if A and B don't have a common eigenvalue.

Solution. I could have written a shorter proof, but I prefer to introduce ideas step by step. First, assume that A and B have a common eigenvalue λ. Then it is also eigenvalue of B^T, since B and B^T have the same characteristic polynomial.

We can find a vector v and a row u such that $Av = \lambda v$ and $uB = \lambda u$ (the latter is equivalent to $B^T u^T = \lambda u^T$).
Take $Y = vu$. Then $AY - YB = Avu - vuB = \lambda vu - \lambda vu = 0$.
So, for $C = 0$ we have an infinite family of solutions kY, and if for a certain C we have at least one solution X, then we also have an infinite family of solutions $X+kY$.

Now consider the case when A and B^T are diagonalizable. So, A has and eigenbasis of vectors $v_1, v_2, \ldots, v_n$ and B^T has eigenbasis of vectors $u_1^T, u_2^T, \ldots, u_n^T$, where u_k are rows. Then $\left\{v_i u_j\right\}$ is an eigenbasis of the operator $X \mapsto AX - XB$.
First let us check that it is a basis in the vector space of $n \times n$ matrixes. Since number of elements equals to a dimension, it is sufficient to show that $\left\{v_i u_j\right\}$ span the space; linear independence will follow. Denote by $\{e_k\}$ the standard basis of $\mathbb{R}^n$. Since both v_i span the space, for any k, we can write $e_k = \sum a_{ki} v_i$. For the same reason, for every m we can write $e_m^T = \sum b_{mj} u_j$. Therefore, for each k and m we have $e_k e_m^T = \sum \sum a_{ki} b_{mj} v_i u_j$. So, matrixes$\left\{v_k u_k\right\}$ span the standard basis for $n \times n$ matrixes (that is, matrixes having 1 in one cell and zero at all other cells) so they really span everything and thus they are basis.
By definition of eigenvector, $Av_i = \lambda_i v_i$, and $u_j B = \mu_j u_j$.
If $X = v_i u_j$, then $AX - BX = Av_i u_j - v_i u_j B = (\lambda_i - \mu_j)X$.
So we see that it is an eigenbasis for the operator $X \mapsto AX - XB$, and its eigenvalues are $\lambda_i - \mu_j$. The operator is invertible iff all eigenvalues are nonzero and that is when all the eigenvalues of A are different from all the eigenvalues of B.

Now, assume that A and B^T are not necessarily invertible. But anyway, for every matrix we can choose a basis (in algebraically closed field) such that the matrix will be upper triangular. In non-coordinate language bringing matrix A to upper triangular form means the following: we can choose a basis $\{v_k\}$, such that Av_k is a linear combination of v_j for $j \leq k$, and the coefficient of v_k in that decomposition is the corresponding eigenvalue.
So, we choose such a basis for matrix A and a basis $\{u_j^T\}$ with the same property for matrix B. Then, for the reasons we've already explained, $\{v_i u_j\}$ is a basis for matrixes. We shall index this basis by $i + nj$ (indexing is needed to define upper triangular property).
Then $Av_i u_j - v_i u_j B = (\lambda_i - \mu_j)\, v_i u_j$ + linear combination of previous basis elements, so this basis brings the operator $X \mapsto AX - XB$ to a triangular form, and $\lambda_i - \mu_j$ appear on the diagonal, QED.

5.* Let A be an invertible $n\times n$ real matrix, U, V be linear subsets of $\mathbb{R}^n$. Assume that U and V are ***almost disjoint***, which means they have no more common elements except 0.
Show that there exists an integer k such that A^kU and V are almost disjoint.

First solution. This solution is very short, but it uses some higher math. I shall try to explain it here, but if it is not clear enough, we shall discuss it in greater detail during one of the meetings. The required higher math in this case is exterior power. If we have a linear space W, we can construct k exterior power of that space, $\wedge^k W$ as follows:

First consider expressions $w_1\wedge w_2\wedge\ldots\wedge w_k$, where $w_i \in$ W.

Consider linear combinations of those expressions. Introduce 3 types of relations:

$$a\,(w_1\wedge w_2\wedge\ldots\wedge w_i\wedge\ldots\wedge w_k)= w_1\wedge w_2\wedge\ldots\wedge(aw_i)\wedge\ldots\wedge w_k$$

$$w_1\wedge w_2\wedge\ldots\wedge(u + w_j)\wedge\ldots\wedge w_k = w_1\wedge w_2\wedge\ldots\wedge u\wedge\ldots\wedge w_k + w_1\wedge w_2\wedge\ldots\wedge w_j\wedge\ldots\wedge w_k$$

$$w_1\wedge w_2\wedge\ldots\wedge w_i\wedge\ldots\wedge w_j\wedge\ldots\wedge w_k = - w_1\wedge w_2\wedge\ldots\wedge w_j\wedge\ldots\wedge w_i\wedge\ldots\wedge w_k$$

The linear space formed by these linear combinations with these relations, is $\wedge^k W$.

The following properties of exterior powers are easy exercises:

1. If dim(W) = n, then $\dim\left(\Lambda^l W\right)=\begin{pmatrix} n \\ l \end{pmatrix}$.
2. There is a natural distributive product: $\Lambda^l W\times\Lambda^m W\to\Lambda^{l+m}W$, (it is called wedge product and denoted by $\wedge$)
3. A linear operator $A:W\to W$ naturally induces a linear operator $A_*:\Lambda^l W\to\Lambda^l W$. If A is invertible, then A_* is also invertible.

Remark. The last fact gives the most generic way to define determinant.
If you solved these exercises, you can read on.

In our problem, we have sub-spaces V and U in $\mathbb{R}^n$.

Let $v_1, v_2, \ldots, v_l$ be a basis of V and $u_1, u_2, \ldots, u_m$ be a basis of U.
Denote $v = v_1\wedge v_2\wedge \ldots \wedge v_l$ and $u = u_1\wedge u_2\wedge \ldots \wedge u_m$.
What we actually need is to find k such that $(A_*^{\,k} u) \wedge v$ is not zero.
A_* is invertible, so by Cayley-Hamilton theorem: $A_*^{\,N}+ k_{N-1}A_*^{\,N-1}+\ldots+ k_1A_* +k_0\mathbf{I} =0$.

Here $\mathbf{I}$ is the identity matrix, $k_0 = \pm\det(A_*) \neq 0$ and $N=\begin{pmatrix} n \\ l \end{pmatrix}$.

Apply this identity to u:

$$A_*^N u + k_{N-1}A_*^{N-1} u +\ldots+ k_1A_* u + k_0u = 0.$$

Multiply it externally by v:

$$(A_*^N u)\wedge v + (k_{N-1}A_*^{N-1} u)\wedge v +\ldots+(k_1A_* u)\wedge v + k_0u\wedge v = 0.$$

The last term in this sum is nonzero, so there must be yet another term in the same sum which is non-zero. QED.

Second solution (Alexey Gladkich). It is sufficient to solve the problem when

$$\dim V + \dim U = n.$$

Otherwise we have a vector v which is not in span of U and V, add it to V and repeat it several times until $\dim V + \dim U = n$.

Let $v_1, v_2, \ldots, v_m$ be a basis of V and $u_1, u_2, \ldots, u_{n-m}$ be a basis of U.
For every k we construct M_k a matrix, first m columns of which are $v_1, v_2, \ldots, v_m$ and last $m - k$ columns are $Au_1, Au_2, \ldots, Au_{n-m}$. The determinant M_k is nonzero iff A^kU and V are almost disjoint.
Also, we may assume that A is of Jordan form in the standard basis.
Recall that the numbers appearing in the k'th power of Jordan cell of eigenvalue λ and size j are $\lambda^k p(k)$ where $p(k)$ is a polynomial of degree less than k (and all eigenvalues are nonzero since the matrix is invertible).
So, the numbers in M_k are sums of expressions of the kind $\lambda^k p(k)$ and so $\det(M_k)$ is sum of products of those so it is also linear combination of expressions of that kind. Then the statement follows to the following theorem, applied to the function $f(k) = \det(M_k)$

Theorem. Consider a function $f(k)$ which is a sum of functions $\lambda_i^k p_i(k)$, where p_i are polynomials, and all $\lambda_i \neq 0$. If $f(0) \neq 1$ then $f(k) \neq 0$ for a positive integer k.

We shall show two proofs for this theorem.

First proof of the theorem. Take terms with highest number $|\lambda_i|$. There can be more than one of that kind.
Out of these, take terms with highest power of x (for example, if you have $10^k k^2$ and $10^k k^3$ take only the last one). These terms after several steps become greater by much than all other terms.
So, the sum of these terms is $k^n\left(a_1\lambda_1^k + \ldots + a_m\lambda_m^k\right) = k^n r^k\left(a_1e^{ikc_1} + \ldots + a_me^{ikc_m}\right)$, where $r = |\lambda_1| = |\lambda_2| = \ldots = |\lambda_m|$.

The bracket $b_k = \left(a_1 e^{ikc_1} + ... + a_m e^{ikc_m}\right)$ can be 0 for all nonnegative integers. In that case, we can easily delete these terms from the sum, and consider a shorter and smaller sum of all the other terms, and do the same thing to it. So, we shall assume that for a certain integer nonnegative *q*, we have $a_1 e^{iqc_1} + ... + a_m e^{iqc_m} = d \neq 0$. Choose $0 < \varepsilon << |d|$. We shall prove that we can find infinite number of *k* as large as we want, such that $|b_k - d| < \varepsilon$. From this it will follow that $|b_k| > |d| - \varepsilon$, hence for those *k*, the expression $k^n r^k \left(a_1 e^{ikc_1} + ... + a_m e^{ikc_m}\right)$ will be growing at least as fast as $k^n r^k \left(|d| - \varepsilon\right)$ and faster than all the other terms in the sum. So absolute values of the whole determinant for those *k* will be very large and far from zero.

The proof of that statement will be based on the following lemma.

Lemma. Given positive real numbers s_1, s_2, … ,s_m, for any $\delta > 0$, we can find infinitely many positive integers *k*, that will be as great as we wish, such that for all *j* the distance from s_j to a positive integer will be less than δ.

First, let us see how this lemma implies the solution of our problem.
Take $s_j = c_j / 2\pi$. If ks_j are close to positive integers, then e^{ikc_1} are close to 1, and b_{k+q} are close to b_k. So to make | $b_{k+q} - b_k$ | < ε , we should simply choose a sufficiently small δ and apply the lemma.
Now it remains only to prove the lemma.

Proof of lemma. It is done by induction over *m*. For *m* = 0 we have an empty set of s_j and element of empty set satisfies every condition.
Suppose we already have a technology to produce sufficiently large *k*'s that satisfy the condition for all numbers except s_m. Let us build a sequence of such numbers, which is very long and each number is bigger by much than the previous, and all numbers satisfy the condition for s_1, …, s_{m-1} with $\delta/2$ instead of δ:

$$k_1, k_2, \ldots, k_N$$

We may assume that $N > 2 + 1/\delta$.
Then $\{k_j s_m\}$ gives us N points on [0,1) interval, and at least two of them, $i < j$ are closer than δ. Then $k = k_j - k_i$ satisfy the condition. QED.

Second proof of the theorem. We shall apply discrete differentiating operators (like in targil 6). We have a function which is $f(k)=\sum_{j=1}^{m}\lambda_j^k p_j(k)$, which is nonzero at 0 and zero at all positive integers.

Consider the operator $\partial_\lambda : f(k)\mapsto g(k)=f(k+1)-\lambda f(k)$.

Applying such an operator to $\lambda_j^k p_j(k)$ (which is an easy exercise) produces $\lambda_j^k q_j(k)$, where q_j is polynomial of the same power as p_j if $\lambda_j \neq \lambda$, and a polynomial of lower power if $\lambda_j = \lambda$.

Also, when we apply such an operator to a function which is nonzero at 0 and 0 at all positive integers, we get again a function of the same kind.

But application of all ∂_{λ_j} sufficiently many times will turn our function into a constant, which is a contradiction.