Summary of week 12 (Lectures 34, 35 and 36)

We have defined the *right null space* of $A \in \mathrm{Mat}(m \times n, F)$ to be the space of all $v \in F^n$ such that $Av = 0$. We call the dimension of the right null space of A the *right nullity* of A. Similarly, the *left null space* is the space of all $v \in {}^{\mathrm{t}}F^m$ such that $vA = 0$, and its dimension is called the *left nullity*.

Recall that A, $B \in \mathrm{Mat}(m \times n, F)$ are said to be *equivalent* if $B = PAQ$ for some invertible P and Q, and that equivalent matrices have the same rank. The Rank-Nullity Theorem (Theorem 7.25 of [VST]) says that the rank of A plus the right nullity of A equals the number of columns of A. Similarly, the rank of A plus the left nullity of A equals the number of rows of A. The proof is a one-line application of the Main Theorem on Linear Transformations. See [VST], p. 166.

The Rank-Nullity Theorem expresses a fact that you should be familiar with concerning the solution space of a homogeneous system of m linear equations in n unknowns. We may write the equations as $Ax = 0$, where $A \in \mathrm{Mat}(m \times n, F)$. To solve, apply elementary row operations to A until an echelon matrix J is obtained. It is easily seen that the nonzero rows of an echelon matrix are necessarily linearly independent, and of course they also span the row space; so the nonzero rows of J form a basis for $\mathrm{RS}(J)$, which equals $\mathrm{RS}(A)$ since elementary row operations do not change the row space. So the number of nonzero rows of J equals the rank of A. Each nonzero row contains a leading entry, and the leading entries all lie in different columns. So the number of free variables, which is the number of columns that do not contain leading entries, is the total number of columns minus the rank. The number of free variables is the dimension of the solution space of the system $Ax = 0$; that is, it is the dimension of the right null space of A. So the right nullity of A is the number of columns of A minus the rank.

Here is an example. The system

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 2 \\ 1 & 1 & 2 & 1 & 3 \end{pmatrix} \begin{pmatrix} v \\ w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

has coefficient matrix of rank 2 (since two obvious elementary row operations eliminate the last row, producing an echelon matrix where the number of nonzero rows is two). There are 5 unknowns; so the solution space must have dimension $5 - 2 = 3$. Indeed, the general solution is

$$\begin{pmatrix} v \\ w \\ x \\ y \\ z \end{pmatrix} = \alpha \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} 1 \\ 0 \\ -2 \\ 0 \\ 1 \end{pmatrix}$$

and the three column vectors appearing on the right hand side form a basis for the solution space.

The remainder of Lecture 34 dealt with §8c of [VST]. First we proved that if A is a $k \times k$ matrix and M and $n \times n$ matrix such that

$$M = \begin{pmatrix} A & X \\ 0 & I \end{pmatrix},$$

where I is the $(n-k) \times (n-k)$ identity matrix, X is a $k \times (n-k)$ matrix and 0 is the $(n-k) \times k$ zero matrix, then $\det(M) = \det(A)$.

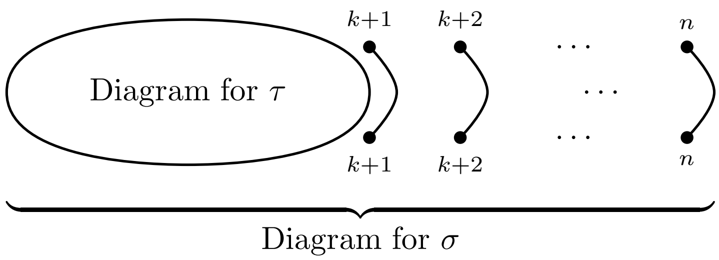
The proof given in lectures proceeded as follows. Let m_{ij} denote the (i,j)-entry of M, and a_{ij} the (i,j)-entry of A. By definition,

$$\det M = \sum_{\sigma \in S_n} \varepsilon(\sigma) m_{1\,\sigma(1)} m_{2\,\sigma(2)} \cdots m_{n\,\sigma(n)}. \tag{1}$$

For all i from $k+1$ to n, the i-th row of M has only one nonzero entry; indeed m_{ij} is 1 for $j = i$ and 0 for $j \neq i$. So if $\sigma \in S_n$ then $m_{1\sigma(1)} m_{2\sigma(2)} \cdots m_{n\sigma(n)}$ is nonzero only if $\sigma(i) = i$ for all i from $k+1$ to n. Thus all nonzero terms on the right hand side of Eq. (1) correspond to permutations $\sigma \in S_n$ of the form

$$\sigma = \begin{bmatrix} 1 & 2 & \ldots & k & k+1 & \ldots & n-1 & n \\ \tau(1) & \tau(2) & \ldots & \tau(k) & k+1 & \ldots & n-1 & n \end{bmatrix}$$

where τ is a permutation of $\{1, 2, \ldots, k\}$. Now it is easily seen that a diagram for τ can be extended to a diagram for σ without changing the number of crossings,



and so it follows that $\varepsilon(\sigma) = \varepsilon(\tau)$. Thus Eq. (1) becomes

$$\det M = \sum_{\tau \in S_k} \varepsilon(\tau) m_{1\,\tau(1)} m_{2\,\tau(2)} \cdots m_{k\,\tau(k)} m_{k+1\,k+1} \cdots m_{nn}.$$

Now since $m_{ij} = a_{ij}$ for all i, $j \in \{1, 2, \ldots, k\}$ and $m_{ii} = 1$ for all $i \in \{k+1, \ldots, n\}$, it follows that

$$\det M = \sum_{\tau \in S_n} \varepsilon(\tau) a_{1\,\tau(1)} a_{2\,\tau(2)} \cdots a_{k\,\tau(k)},$$

and this is exactly $\det A$, as required.

Suppose now that M is an $n \times n$ matrix of the form

$$M = \begin{pmatrix} A_{11} & \square & A_{12} \\ 0 \dots 0 & 1 & 0 \dots 0 \\ A_{21} & \square & A_{22} \end{pmatrix}$$

where the 1 is in the (i, j) position, the rectangles indicate unspecified entries in the j-th column, and the A_{rs} are matrices of the appropriate sizes. By a total of $n - i$ row-swapping and $n - j$ column-swapping operations we can convert M to the form

$$\begin{pmatrix} A_{11} & A_{12} & \square \\ A_{21} & A_{22} & \square \\ 0\,0 \dots\dots & 0\,0 & 1 \end{pmatrix}$$

Since each row swap and each column swap multiply the determinant by -1, we conclude that

$$\det M = (-1)^{n-i}(-1)^{n-j} \det \begin{pmatrix} A_{11} & A_{12} & \square \\ A_{21} & A_{22} & \square \\ 0\,0 \dots\dots & 0\,0 & 1 \end{pmatrix} = (-1)^{i+j} \det \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \tag{2}$$

by the result proved above (applied in the case $k = 1$). Observe that the last matrix in the equation above is the matrix obtained from the original matrix M by deleting the i-th row and the j-th column.

Definition. If A is an $n \times n$ matrix then for all i, $j \in \{1, 2, \dots, n\}$ we define the (i, j)-*cofactor* of A to be $(-1)^{i+j}$ times the determinant of the matrix obtained from A by deleting row i and column j. We denote the (i, j)-cofactor of A by $\mathrm{cof}_{ij}(A)$.

Now let A be any $n \times n$ matrix, and let the i-th row of A be $(a_{i1}\ a_{i2}\ \dots\ a_{in})$. Since

$$(a_{i1}\ a_{i2}\ \dots\ a_{in}) = a_{i1}(1\ 0\ 0\ \dots\ 0) + a_{i2}(0\ 1\ 0\ \dots\ 0) + \cdots + a_{in}(0\ 0\ 0\ \dots\ 1),$$

linearity of the determinant as a function of the i-th row yields that $\det A$ is

$$\det \begin{pmatrix} X \\ a_{i1} \cdots a_{in} \\ Y \end{pmatrix} = a_{i1} \det \begin{pmatrix} X \\ 1\,0 \dots 0 \\ Y \end{pmatrix} + a_{i2} \det \begin{pmatrix} X \\ 0\,1 \dots 0 \\ Y \end{pmatrix} + \cdots + a_{in} \det \begin{pmatrix} X \\ 0\,0 \dots 1 \\ Y \end{pmatrix}$$

where X and Y designate, respectively, the first $i - 1$ rows and the last $n - i$ rows of A. But now deleting row i and column j from the matrix

$$\begin{pmatrix} X \\ 0..1..0 \\ Y \end{pmatrix}$$

(where the 1 is in column j) gives the same result as deleting row i and column j from A; so in view of the formula in Eq. (2) above it follows that

$$\det A = \sum_{j=1}^{n} a_{ij} \mathrm{cof}_{ij}(A).$$

This is known as the *i-th row expansion formula* for the determinant of A.

Since transposing a matrix changes rows into columns, but does not change the determinant, it follows that there is also a *j-th column expansion formula*: $\det A = \sum_{i=1}^{n} a_{ij}\mathrm{cof}_{ij}(A)$.

Definition. If A is an $n \times n$ matrix then the *adjoint matrix*, $\operatorname{adj} A$, is defined to be the transposed matrix of cofactors; that is, the (i, j)-entry of $\operatorname{adj} A$ is $\mathrm{cof}_{ji}(A)$, for all i, $j \in \{1, 2, \ldots, n\}$.

It is a consequence of the i-th row expansion formula for the determinant that $A(\operatorname{adj} A) = (\det A)I$. See [VST] (Theorem 8.25) for the proof. A similar proof, using column expansions instead of row expansions, gives $(\operatorname{adj} A)A = (\det A)I$. It follows that if $\det A \neq 0$ then

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A.$$

This formula is important, but not recommended as a means of calculating inverses.

Lectures 35 and 36 were concerned with the material §9a, §9b and §9d of [VST]. For §9a the treatment given in lectures was similar to that in the book. (Note that a matrix version of Theorem 9.6 was proved in an earlier lecture.)

If $\phi\colon V \to V$ is a linear operator and U is a subspace of V then U is said to be *ϕ-invariant* if $\phi(x) \in U$ for all $x \in U$. (This is Definition 9.8 of [VST].) Given a ϕ-invariant subspace U we can define a linear operator $\phi_U\colon U \to U$ by the rule that $\phi_U(x) = \phi(x)$ for all $x \in U$. The operator ϕ_U is called the *restriction* of ϕ to U. Note that although $\phi_U(x) = \phi(x)$ for all $x \in U$, the functions ϕ_U and ϕ are not the same (unless $U = V$), since their domains and codomains are different: if $v \in V$ and $v \notin U$ then $\phi(v)$ is defined but $\phi_U(v)$ is not. The relationship between ϕ_U and ϕ can be depicted diagrammatically as below. The map from $U \to V$ indicated by the vertical hooked arrows is the so-called *inclusion map*, which takes each $u \in U$ to itself considered as an element of V. The two alternative ways to follow the arrows from the lower left corner to the upper right corner give the same result.

$$\begin{tikzcd} V \arrow[r, "\phi"] & V \\ U \arrow[u, hook] \arrow[r, "\phi_U"'] & U \arrow[u, hook] \end{tikzcd}$$

If ϕ is a linear operator on a vector space V, and V can be expressed as the direct sum of two ϕ-invariant subspaces U and W, then the matrix of ϕ, relative to a basis of V made by combining a basis of U and a basis of W, has a particularly simple form. See Theorem 9.9 of [VST]. It follows from Theorem 9.9 that in this situation the characteristic polynomial of ϕ is the product of the characteristic polynomials of ϕ_U and ϕ_W. This is Theorem 9.10 of [VST]. Note that the paragraph on the lower half of p. 203 can be ignored: quotient spaces are not part of the course.

Definition. A linear operator ϕ on a vector space V is said to be *nilpotent* if some power of ϕ is the zero map. That is, ϕ is nilpotent if there exists a positive integer n such that $\phi^n(v) = 0$ for all $v \in V$.

Let ϕ be a linear operator on the space V, and assume that V is finite-dimensional. For each positive integer i define K_i to be the kernel of ϕ^i. That is,

$$K_i = \{\, v \in V \mid \phi^i(v) = 0 \,\}.$$

If $v \in K_i$ then $\phi^{i+1}(v) = \phi(\phi^i(v)) = \phi(0) = 0$, and so $v \in K_{i+1}$. Since this holds for all $v \in K_i$ it follows that $K_i \subseteq K_{i+1}$. Hence the K_i form an increasing chain of subspaces of V:

$$\{0\} \subseteq K_1 \subseteq K_2 \subseteq K_3 \subseteq \cdots\cdots\cdots \subseteq V.$$

It follows that if $d_i = \dim K_i$ then the d_i form an increasing† sequence of non-negative integers, all of which are less than or equal to $\dim V$. Since a bounded set of integers must have a maximal element, there exists an integer r such that $d_r \geq d_i$ for all i, and it then follows that $d_r = d_i$ for all $i \geq r$:

$$0 \leq d_1 \leq d_2 \leq d_3 \leq \cdots \leq d_r = d_{r+1} = d_{r+2} = \cdots .$$

For $i \geq r$ we also have that $K_i = K_r$, since $K_r \subseteq K_i$ and $\dim K_r = \dim K_i$ (by Proposition 4.11 of [VST]). The chain of subspaces K_i reaches a maximum and stays there:

$$\{0\} \subseteq K_1 \subseteq K_2 \subseteq K_3 \subseteq \cdots \subseteq K_r = K_{r+1} = K_{r+2} = \cdots .$$

Observe that K_r is the set of all $v \in V$ that are annihilated by some power of ϕ:

$$K_r = \{\, v \in V \mid \phi^i(v) = 0 \text{ for some } i \in \mathbb{Z} \,\}.$$

We call this set the *generalized kernel* of ϕ. Observe that ϕ is nilpotent if and only if the generalized kernel is the whole space V.

Continuing with the same notation as above, let H_i be the image of the operator ϕ^i. In particular, with r as above,

$$H_r = \{\, v \in V \mid v = \phi^r(u) \text{ for some } u \in V \,\}.$$

---

† Perhaps I should say "non-decreasing" rather than "increasing", since the inequalities need not be strict.

Observe that, by The Main Theorem on Linear Transformations,

$$\dim K_r + \dim H_r = \dim V. \tag{3}$$

Now suppose that $v \in K_r \cap H_r$. Then $v = \phi^r(u)$ for some $u \in V$, since $v \in H_r$, and since $v \in K_r$ we find that

$$0 = \phi^r(v) = \phi^r(\phi^r(u)) = \phi^{2r}(u).$$

Hence $u \in \ker \phi^{2r} = K_{2r}$, which equals K_r since we proved above that $K_i = K_r$ for all $i \geq r$. But $u \in K_r$ means that $\phi^r(u) = 0$, and since u was chosen so that $v = \phi^r(u)$ we conclude that $v = 0$. But since v was an arbitrary element of $K_r \cap H_r$ we deduce that 0 is the only element of $K_r \cap H_r$. And since $H_r \cap K_r = \{0\}$, the sum of the subspaces K_r and H_r is a direct sum: $K_r + H_r = K_r \oplus H_r$.

By Theorem 6.9 of [VST], the dimension of a direct sum is the sum of the dimensions of the direct summands; so

$$\dim(K_r \oplus H_r) = \dim K_r + \dim H_r = \dim V,$$

by Eq. (3) above, and since $K_r \oplus H_r$ is a subspace of V it follows that $K_r \oplus H_r = V$. It is easy to show—although the proof was omitted—that the subspaces H_r and K_r are ϕ-invariant. So, in effect, ϕ is made up of two component operators, ϕ_{K_r} (acting on the direct summand K_r) and ϕ_{H_r} (acting on the direct summand H_r). By Theorem 9.9, the characteristic polynomial of ϕ is the product of the characteristic polynomials of ϕ_{K_r} and ϕ_{H_r}.

Since K_r is the generalized kernel of ϕ we see that ϕ_{K_r} is nilpotent. Let us now show that ϕ_{H_r} is bijective (and therefore invertible). If $x \in \ker \phi_{H_r}$ then $x \in H_r$ and $\phi(x) = 0$; so

$$x \in H_r \cap (\ker \phi) = H_r \cap K_1 \subseteq H_r \cap K_r = \{0\},$$

and therefore $\ker \phi_{H_r} = \{0\}$. It follows by Proposition 3.15 of [VST] that ϕ_{H_r} is injective. But the Main Theorem on Linear Transformations implies that an injective linear map from a finite-dimensional space to another of the same dimension is necessarily also surjective (see Tutorial 11); so ϕ_{H_r} is both injective and surjective, and therefore bijective. Thus we have shown that an arbitrary linear operator on a finite-dimensional space is, in some sense, made up of a nilpotent piece and an invertible piece.

An invertible linear operator ψ does not have zero as an eigenvalue: if 0 were an eigenvalue then there would be a nonzero v with $\psi(v) = 0v = 0$.† This means that $\phi(v) = \phi(0)$, contradicting the fact that an invertible map must be injective.

† Note that the first 0 here is the zero scalar, the other the zero vector.

By contrast, a nilpotent operator θ on a space U can have no eigenvalues other than 0. For suppose that λ is an eigenvalue of θ. Then $\theta(v) = \lambda v$ for some nonzero v, and now we find, successively, that

$$\begin{aligned}\theta^2(v) &= \theta(\theta(v)) = \theta(\lambda v) = \lambda\theta(v) = \lambda^2 v,\\ \theta^3(v) &= \theta(\theta^2(v)) = \theta(\lambda^2 v) = \lambda^2\theta(v) = \lambda^3 v,\\ \theta^4(v) &= \theta(\theta^3(v)) = \theta(\lambda^3 v) = \lambda^3\theta(v) = \lambda^4 v,\end{aligned}$$

and so on. But the nilpotence of θ implies that $\theta^n(v) = 0$ for some n; so $\lambda^n v = 0$, and since $v \neq 0$ this implies (by Proposition 3.7) that $\lambda^n = 0$ and hence $\lambda = 0$.

Since θ has no eigenvalues other than zero, the characteristic polynomial $c_\theta(x)$ must be a scalar multiple of x^d, where $d = \dim U$. This can also be seen by calculating the matrix of θ relative to a suitable basis. If K_l is the kernel of θ^l then the subspaces K_l form an increasing chain, with $K_r = U$ for some r. We can choose a basis of K_1, extend this basis to a basis of K_2, extend again to a basis of K_3, and continue in this way until we have a basis $u_1, u_2, \ldots, u_d$ of U that includes bases of all the subspaces K_l (for l from 0 to r). That is, if $d_l = \dim K_l$ then the vectors $u_1, u_2, \ldots, u_{d_l}$ form a basis of K_l.

Let M be the matrix of θ relative to the basis $u_1, u_2, \ldots, u_d$ of U, so that for all j,

$$\theta(u_j) = \sum_{i=1}^{d} \alpha_{ij} u_i \tag{4}$$

where α_{ij} is the (i,j)-entry of M. For any j we may choose the least l such that $\theta^l(u_j) = 0$. Then $u_j \in K_l$ and $u_j \notin K_{l-1}$, and it follows that $d_{l-1} < j \leq d_l$. Since $\theta^{l-1}(\theta(u_j)) = 0$ we see that $\theta(u_j) \in \ker\theta^{l-1} = K_{l-1}$, and so $\theta(u_j)$ is a linear combination of $u_1, u_2, \ldots, u_{d_{l-1}}$. Since these precede u_j in the sequence $u_1, u_2, \ldots, u_d$, the scalars α_{ij} in Eq. (4) are zero whenever $i \geq j$. This means that in the matrix M all the entries on and below the leading diagonal are zero. Thus the characteristic polynomial $c_\theta(x)$ (which is the same as the characteristic polynomial of M) is

$$\det(M - xI) = \det\begin{pmatrix} -x & * & * & \ldots & * \\ 0 & -x & * & \ldots & * \\ 0 & 0 & -x & \ldots & * \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & -x \end{pmatrix} = (-x)^d$$

(where the $*$'s denote the various above diagonal entries of M).

Now consider once again an arbitrary linear operator ϕ on a finite-dimensional space V, and let λ be an eigenvalue of ϕ, so that $\lambda - x$ is a factor of the characteristic polynomial $c_\phi(x)$. Write $c_\phi(x) = (\lambda - x)^m q(x)$, where $q(x)$ does not have $\lambda - x$

as a factor. (Thus m is the multiplicity of $\lambda - x$ as a factor of $c_\phi(x)$.) If we put $\psi = \phi - \lambda\mathrm{i}$, where i is the identity operator on V, then

$$c_\psi(x) = \det(\psi - x\mathrm{i}) = \det(\phi - (\lambda + x)\mathrm{i}) = c_\phi(x + \lambda).$$

The λ-eigenspace of ϕ is $\ker(\phi - \lambda\mathrm{i}) = \ker\psi$, and we define the *generalized λ-eigenspace of ϕ* to be the generalized kernel of ψ: the set

$$G_\lambda = \{\, v \in V \mid (\phi - \lambda\mathrm{i})^l(v) = 0 \text{ for some } l \in \mathbb{Z}\,\}.$$

As shown above, we have that $V = G_\lambda \oplus H$ for some subspace H that is ψ-invariant and therefore also ϕ-invariant. Observe that $\phi_{G_\lambda} - \lambda\mathrm{i} = \psi_{G_\lambda}$, and $\phi_H - \lambda\mathrm{i} = \psi_H$. If we write $f(x)$ and $g(x)$ for the characteristic polynomials of ϕ_{G_λ} and ϕ_H respectively, so that $f(x)g(x) = c_\phi(x)$ by Theorem 9.9, then $f(x+\lambda)$ and $g(x+\lambda)$ are the characteristic polynomials of ψ_{G_λ} and ψ_H. But ψ_{G_λ} is nilpotent; so its characteristic polynomial is $(-x)^{d_\lambda}$, where $d_\lambda = \dim G_\lambda$; so $f(x+\lambda) = (-x)^{d_\lambda}$, giving

$$f(x) = (\lambda - x)^{d_\lambda}.$$

Furthermore, we also know that 0 is not an eigenvalue of ψ_H, and so x is not a factor of $g(x + \lambda)$. So $x - \lambda$ is not a factor of $g(x)$, and since

$$c_\phi(x) = f(x)g(x) = (\lambda - x)^{d_\lambda} g(x)$$

we conclude that the multiplicity of $\lambda - x$ as a factor of $c_\phi(x)$ is the dimension of the generalized λ-eigenspace of ϕ. That is, $d_\lambda = m$ and $g(x) = q(x)$, in the notation used above.

If we now choose another eigenvalue μ of ϕ then the $\mu - x$ must be a factor of $g(x) = c_{\phi_H}(x)$, and, indeed, its multiplicity as a factor of $g(x)$ is the same as its multiplicity in $c_\phi(x)$. We can repeat the above reasoning with ϕ replaced by ϕ_H and λ by μ, and deduce that $H = G_\mu \oplus H'$, giving $V = G_\lambda \oplus G_\mu \oplus H'$, where G_μ is the generalized μ-eigenspace of ϕ_H and H' is some ϕ_H-invariant subspace. Moreover, we see that $\dim G_\mu$ is the multiplicity of $\mu - x$ as a factor of $c_\theta(x)$, and it is not hard to deduce from this that G_μ coincides with the generalized μ-eigenspace of the operator ϕ on the whole space V. Provided that the field we are working over is algebraically closed we can keep repeating this process. So long as the space H'' is nonzero the characteristic polynomial of the restriction of ϕ to H'' will have some factor $\nu - x$, and then $H' = G_\nu \oplus H''$ for some ϕ-invariant H'', where G_ν is the generalized ν-eigenspace. This gives $V = G_\lambda \oplus G_\mu \oplus G_\nu \oplus H''$. The final summand gets smaller with each step, and so we obtain eventually that V is the direct sum of the generalized eigenspaces corresponding to all of its eigenvalues,

$$V = G_{\lambda_1} \oplus G_{\lambda_2} \oplus \cdots \oplus G_{\lambda_s},$$

the characteristic polynomial of the restriction of ϕ to G_{λ_i} being $(\lambda_i - x)^{m_i}$, where

$$c_\phi(x) = (\lambda_1 - x)^{m_1}(\lambda_2 - x)^{m_2} \cdots (\lambda_s - x)^{m_s}$$

and λ_1, λ_2, …, λ_s are pairwise distinct.