The Decomposition Theorem

The aim of this section is to prove the following theorem

Theorem 1 (Decomposition Theorem). *Let V be a vector space over a field K and let T be a linear operator on V. If $a_1, a_2, ..., a_m$ are distinct scalars and $k_1, k_2, ..., k_m \in \mathbb{N}$ then*

$$\mathrm{Ker}((\mathrm{T}-\mathrm{a}_1)^{\mathrm{k}_1}(\mathrm{T}-\mathrm{a}^2)^{\mathrm{k}_2}\cdots(\mathrm{T}-\mathrm{a}_\mathrm{m})^{\mathrm{k}_\mathrm{m}}) = \mathrm{Ker}((\mathrm{T}-\mathrm{a}_1)^{\mathrm{k}_1})\oplus\mathrm{Ker}((\mathrm{T}-\mathrm{a}_2)^{\mathrm{k}_2})\oplus\cdots\oplus\mathrm{Ker}((\mathrm{T}-\mathrm{a}_\mathrm{m})^{\mathrm{k}_\mathrm{m}}).$$

Corollary 2. *A linear operator on a finite-dimensional vector space is diagonalizable iff its minimal polynomial is a product of distinct linear factors.*

Lemma 3. *Let T be a linear operator with $T^k = 0$, $k \geq 1$. Then $1 - T$ is invertible with*

$$(1-T)^{-1} = 1 + T + T^2 + \cdots + T^{k-1}.$$

Proof. $(1-T)(1+T+\cdots+T^{k-1}) = 1+T+\cdots+T^{k-1}-T-\cdots-T^k = 1$. This proves the result since any two polynomials in T commute. □

Corollary 4. *If T is a linear operator with $(T-a)^k = 0$ for some $k \geq 1$ and some scalar a then, for any scalar $c \neq a$, the operator $T - c$ is invertible with*

$$(c-T)^{-1} = (c-a)^{-1} + (c-a)^{-2}(T-a) + \cdots + (c-a)^{-k}(T-a)^{k-1}.$$

Proof. We have $c - T = c - a - (T-a) = (c-a)(1-(T-a)/(c-a))$. □

Proof of Decomposition Theorem. Let $V_i = \ker(T-a)^{k_i}$. We first show that the sum $V_1+V_2+\cdots+V_m$ is direct. If $u \in V_i$ we must have $T(u) \in V_i$ since

$$(T-a)^{k_i}(T(u)) = (T-a)^{k_i}T(u) = T(T-a)^{k_i}(u) = T((T-a)^{k_i}(u)) = 0.$$

If R_i is the restriction of T to V_i then R_i is a linear operator on V_i with $(R_i - a_i)^{k_i} = 0$. It follows that $R_i - c$ is invertible for any scalar $c \neq a_i$. Now let $v_i \in V_i$ for $1 \leq i \leq m$ and suppose that $v_1 + v_2 + \cdots + v_m = 0$. We want to show that $v_i = 0$. But

$$0 = \prod_{j \neq i}(T-a_j)^{k_j}(v_1+v_2+\cdots+v_m) = S_i(v_i)$$

with $S_i = \prod_{j\neq i}(R_i - a_j)^{k_j}$, an invertible operator on V_i. It follows that $v_i = 0$ for all i and so the sum is direct.

To show that the sum is $\ker((T-a_1)^{k_1}(T-a^2)^{k_2}\cdots(T-a_m)^{k_m})$ we proceed by induction on m. Assume that the result is true for some $m = p$ and let

$$v \in \mathrm{Ker}((\mathrm{T}-\mathrm{a}_1)^{\mathrm{k}_1}(\mathrm{T}-\mathrm{a}^2)^{\mathrm{k}_2}\cdots(\mathrm{T}-\mathrm{a}_{\mathrm{p}+1})^{\mathrm{k}_{\mathrm{p}+1}}).$$

Then $w = (T-a_{p+1})^{k_{p+1}}(v) \in \mathrm{Ker}((\mathrm{T}-\mathrm{a}_1)^{\mathrm{k}_1}(\mathrm{T}-\mathrm{a}^2)^{\mathrm{k}_2}\cdots(\mathrm{T}-\mathrm{a}_\mathrm{p})^{\mathrm{k}_\mathrm{p}})$. By our inductive hypothesis we have $w = w_1 + w_2 + \cdots + w_p$ with $w_i \in V_i$. But, by the invertibility of the restriction of $T - a_{p+1}$ to V_i for $1 \leq i \leq p$, we have $w_i = (T-a_i)^{k_i}(v_i)$ with $v_i \in V_i$. Hence

$$(T-a_{p+1})^{k_{p+1}}(v) = (T-a_{p+1})^{k_{p+1}}(v_1+v_2+\cdots+v_p)$$

which shows that $v_{p+1} = v - (v_1+v_2+\cdots+v_p) \in \mathrm{Ker}(\mathrm{T}-\mathrm{a}_{\mathrm{p}+1}$. Thus $v = v_1 + v_2 + \cdots + v_{p+1}$ yielding the result for $m = p+1$. □

Example 1. *A function $y = f(x)$ is a solution of the homogeneous differential equation*

$$y''' - 4y'' + 5y' - 2y = 0$$

if and only iff f is infinitely differentiable and

$$f \in \mathrm{Ker}(\mathrm{D}^3 - 4\mathrm{D}^2 + 5\mathrm{D} - 2) = \mathrm{Ker}((\mathrm{D}-1)^2(\mathrm{D}-2)) = \mathrm{Ker}((\mathrm{D}-1)^2) \oplus \mathrm{Ker}(\mathrm{D}-2),$$

where D is the differentiation operator on $C^\infty(\mathbb{R})$, the vector space of infinitely differentiable real valued functions on $\mathbb{R}$. Since $\mathrm{Ker}((\mathrm{D}-1)^2) = \mathrm{Span}(\mathrm{e^x}, \mathrm{xe^x})$ and $\mathrm{Ker}(\mathrm{D}-2) = \mathrm{Span}(\mathrm{e^{2x}})$, this shows this the solutions of the given diiferential equation are the functions

$$y = f(x) = ae^x + bxe^x + ce^2x.$$

To solve the non-homogeneous differential equation

$$y''' - 4y'' + 5y' - 2y = e^{3x}$$

it suffices to find one solution y_P since, for any other solution y, the function $y - y_P$ is a solution of the associated homogeneous system; so $y = y_P + ae^x + bxe^x + ce^{2x}$ is the general solution of the given non-homogeneous equation. Since $e^{3x} \in \mathrm{Ker}(\mathrm{D}-3)$ and the restrictions of $D-1$, $D-2$ to $\mathrm{Ker}(\mathrm{D}-3)$ are invertible, we must have $y_P = Ce^{3x}$. Substituting this in the given equation, we get $C = 1/4$.
If we want to find a particular solution of $y''' - 4y'' + 5y' - 2y = e^x$, one has to proceed differently as $(D-1)$ is not invertible on $\mathrm{Ker}(\mathrm{D}-1)$. Since $y_P \in \mathrm{Ker}((\mathrm{D}-1)^3(\mathrm{D}-2))$, we have $y_P = Cx^2e^x$. Substituting this in the given differential equation, we get $C = -1/2$.

Example 2. *To find a formula for $s_n = \sum_{i=0}^{i=n}(i^2-1)2^i$ we use the fact that*

$$s_{n+1} - s_n = ((n+1)^2 - 1)2^{n+1} = (2n^2 + 4n)2^n$$

to deduce that the infinite sequence $s = (s_0, s_1, ..., s_n, ...)$ is in the kernel of $(L-1)(L-2)^3$, where L is the left-shift operator on $\mathbb{R}^\infty$. But, using the fact that

$$\mathrm{Ker}((\mathrm{L}-\mathrm{r})^\mathrm{k}) = \mathrm{Span}((1), (\mathrm{r^n}), \ldots, (\mathrm{n^{k-1}r^n})),$$

we obtain

$$\mathrm{Ker}((\mathrm{L}-1)(\mathrm{L}-2)^3) = \mathrm{Ker}(\mathrm{L}-1) + \mathrm{Ker}((\mathrm{L}-2)^3) = \mathrm{Span}((1), (2^\mathrm{n}), (\mathrm{n}2^\mathrm{n}), (\mathrm{n}^2 2^\mathrm{n})),$$

which shows that there must be scalars a, b, c, d such that

$$s_n = a + b2^n + cn2^n + dn^2 2^n$$

for $n \geq 0$. But $s_0 = -1, s_1 = -1, s_2 = 10, s_3 = 74$ and so a, b, c, d is a solution of the equations

$$\begin{aligned} a + b &= -1 \\ a + 2b + 2c + 2d &= -1 \\ a + 4b + 8c + 16d &= 10 \\ a + 8b + 24c + 72d &= 74 \end{aligned}$$

which has the unique solution $a = -10, b = 9, c = -7, d = 5/2$. (Can you explain why a priori this system would have a unique solution?) Thus $s_n = -10 + 9 \cdot 2^n - 7n2^n - 5n^2 2^{n-1}$ for $n \geq 0$.

Let V be a vector space over a field K and let T be a linear operator on V. A subspace W is said to be T-*invariant* if $T(W) \subseteq W$. If $f_1, f_2, ..., f_n$ is a basis of V with $f_1, ..., f_m$ a basis of W then $A = [T]_f = [a_{ij}]$ has a block decomposition

$$\begin{bmatrix} B & C \\ D & E \end{bmatrix}$$

with $B = [a_{ij}]_{1 \le i,j \le m}$, $C = [a_{ij}]_{1 \le i \le m, j > m}$, $D = [a_{ij}]_{i > m, 1 \le j \le m}$, $E = [a_{ij}]_{i,j > m}$. We have $D = 0$ iff W is T-invariant and $C = 0$ iff $\mathrm{Span(f_{m+1}, ..., f_n)}$ is T-invariant. It follows by induction that V is a direct sum of finitely many T-invariant subspaces iff T has a matrix representation which is in block-diagonal form.
If S and T are linear operators on V which commute, i.e. $ST = TS$, then $\mathrm{Ker(S)}$ and $\mathrm{Im(S)}$ are T-invariant. This is left as an execise for the reader. If V is a direct sum of T-invariant subspaces, then T can be represented by a block-diagonal matrix.

Example 3. *Let D be the differentiation operator on $C^\infty(\mathbb{R})$ and let*

$$V = \mathrm{Ker((D - a_1)^{k_1}(D - a_2)^{k_2} \cdots (D - a_m)^{k_m})}$$

with $a_1, ..., a_m$ distinct scalars. Then V is the direct sum of the D-invariant subspaces $V_i = \mathrm{Ker(D - a_i)^{k_i}}$. Moreover, with respect to the basis

$$e^{a_i x}, xe^{a_i x}, \frac{1}{2}x^2 e^{a_i x}, \ldots, \frac{1}{(k_m - 1)!} x^{k_m - 1} e^{a_i x}$$

of V_i, the matrix of the restriction of D to V_i is the $k \times k$ matrix

$$J_k(a) = \begin{bmatrix} a & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & a & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & a & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & a & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & a & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & a \end{bmatrix}$$

with $a = a_i$, $k = k_i = \mathrm{dim}(V_i)$. The matrix $J_k(a)$ is called a Jordan matrix of size k and diagonal element a. The minimal polynomial of this matrix is $(\lambda - a)^k$ which is also equal to the characteristic polynomial. The restriction of D to V thus has a block-diagonal representation with m Jordan blocks, the i-th being $J_{k_i}(a_i)$, and its characteristic and minimal polynomials are both equal to

$$(\lambda - a_1)^{k_1}(\lambda - a_2)^{k_2} \cdots (\lambda - a_m)^{k_m}.$$

In the next section we will show that any linear operator on a finite-dimensional vector space whose minimal polynomial is a product of linear factors has a matrix representation in block-diagonal form with Jordan blocks. Since any polynomial over $\mathbb{C}$ is a product of linear factors, any linear operator on a finite- dimenasional vector space over $\mathbb{C}$ has such a block-diagonal form. In particular, any matrix over $\mathbb{C}$ is similar to one in block-diagonal form with Jordan blocks. Moreover, we shall show that these Jordan blocks are unique up to a permuation of the blocks. This is the so-called Jordan canonical form.