Eigenvalues, diagonalization, and Jordan normal form

Zdeněk Dvořák

April 20, 2016

Definition 1. *Let A be a square matrix whose entries are complex numbers. If $Av = \lambda v$ for a complex number λ and a non-zero vector v, then λ is an eigenvalue of A, and v is the corresponding eigenvector.*

Definition 2. *Let A be a square matrix. Then*

$$p(x) = \det(A - Ix)$$

is the characteristic polynomial of A.

1 Matrix similarity

Definition 3. *Square matrices A and D are similar if $A = CDC^{-1}$ for some regular matrix C. Equivalently, they are similar if they are matrices of the same linear function, with respect to different bases.*

Lemma 1. *If A and D are similar, then they have the same characteristic polynomials, and thus they have the same eigenvalues with the same algebraic multiplicities. Furthermore, their eigenvalues also have the same geometric multiplicities.*

Proof. Suppose that $A = CDC^{-1}$. Then the characteristic polynomial $\det(A - Ix)$ of A is equal to

$$\begin{aligned}\det(A - Ix) &= \det(CDC^{-1} - Ix) = \det(C(D - Ix)C^{-1}) \\ &= \det(C)\det(D - Ix)\det(C^{-1}) = \det(D - Ix),\end{aligned}$$

which is the characteristic polynomial of D.

Furthermore, for any λ, we have $v \in \mathrm{Ker}(D - \lambda I)$ if and only if $Cv \in \mathrm{Ker}(A - \lambda I)$, and since C is regular, we have $\dim(\mathrm{Ker}(D - \lambda I)) = \dim(\mathrm{Ker}(A - \lambda I))$; hence, the geometric multiplicities of λ as an eigenvalue of A and D coincide. □

Corollary 2. *If A and B are square matrices, then AB and BA have the same eigenvalues.*

Proof. Let $X = \begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix}$ and $Y = \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix}$. Note that AB and X have the same eigenvalues except for extra zero eigenvalues of X, and that BA and Y have the same eigenvalues except for extra zero eigenvalues of Y. Furthermore, let $C = \begin{pmatrix} I & A \\ 0 & I \end{pmatrix}$, and note that $XC = CY$, and thus X and Y are similar and have the same eigenvalues by Lemma 1. □

Observation 3. *If $A = CDC^{-1}$ for some square matrices A and D, then $A^n = CD^nC^{-1}$. More generally, for any polynomial p, we have $p(A) = Cp(D)C^{-1}$.*

2 Diagonalization

Example 1. *Let $A = \begin{pmatrix} -2 & -1 & -2 \\ 4 & 3 & 2 \\ 5 & 1 & 5 \end{pmatrix}$ and let $f : \mathbf{R}^3 \to \mathbf{R}^3$ be defined by $f(x) = Ax$.*

Eigenvectors and eigenvalues of A are

- $v_1 = (1, -1, -1)^T$, *eigenvalue 1,*
- $v_2 = (-1, 2, 1)^T$, *eigenvalue 2,*
- $v_3 = (-1, 1, 2)^T$, *eigenvalue 3.*

Note that $B = v_1, v_2, v_3$ is a basis of $\mathbf{R}^3$. If $[x]_B = (\alpha_1, \alpha_2, \alpha_3)$, then

$$\begin{aligned} f(x) &= f(\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3) \\ &= \alpha_1 f(v_1) + \alpha_2 f(v_2) + \alpha_3 f(v_3) \\ &= \alpha_1 v_1 + 2\alpha_2 v_2 + 3\alpha_3 v_3, \end{aligned}$$

and thus $[f(x)]_{B,B} = (\alpha_1, 2\alpha_2, 3\alpha_3)$. Therefore, the matrix of f with respect to the basis B is

$$[f]_{B,B} = D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Let B' be the standard basis of $\mathbf{R}^3$. Let

$$C = [id]_{B,B'} = (v_1|v_2|v_3) = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}.$$

Recall that

$$[f]_{B',B'} = [id]_{B,B'}[f]_{B,B}[id]_{B',B} = [id]_{B,B'}[f]_{B,B}[id]_{B,B'}^{-1},$$

and thus

$$A = CDC^{-1}.$$

Lemma 4. *An $n \times n$ matrix A is similar to a diagonal matrix if and only if there exists a basis of $\mathbf{C}^n$ formed by eigenvectors of A.*

Proof. Suppose that $A = CDC^{-1}$ for a diagonal matrix D with diagonal entries $\lambda_1, \ldots, \lambda_n$. Since C is regular, $B = Ce_1, \ldots, Ce_n$ is a basis of $\mathbf{C}^n$. Furthermore, $A(Ce_i) = CDe_i = \lambda_i(Ce_i)$, and thus B is formed by eigenvectors of A.

Conversely, suppose that $v_1, \ldots, v_n$ is a basis of $\mathbf{C}^n$ formed by eigenvectors of A corresponding to eigenvalues $\lambda_1, \ldots, \lambda_n$, and let D be the diagonal matrix D with diagonal entries $\lambda_1, \ldots, \lambda_n$. Let $C = (v_1|\ldots|v_n)$. Then

$$C^{-1}AC = C^{-1}(Av_1|\ldots|Av_n) = C^{-1}(\lambda_1 v_1|\ldots|\lambda_n v_n) = C^{-1}CD = D,$$

and thus A and D are similar. □

Lemma 5. *If $\lambda_1, \ldots, \lambda_k$ are pairwise distinct eigenvalues of A (not necessarily all of them) and $v_1, \ldots, v_k$ are corresponding eigenvectors, then $v_1, \ldots, v_k$ are linearly independent.*

Proof. We proceed by induction on k; the claim is trivial for $k = 1$. Suppose that $\alpha_1 v_1 + \ldots + \alpha_k v_k = o$; then $o = A(\alpha_1 v_1 + \ldots + \alpha_k v_k) = \alpha_1\lambda_1 v_1 + \ldots + \alpha_k\lambda_k v_k$, and $\alpha_1(\lambda_1 - \lambda_k)v_1 + \alpha_2(\lambda_2 - \lambda_k)v_2 + \ldots + \alpha_{k-1}(\lambda_{k-1} - \lambda_k)v_{k-1} = o$. By the induction hypothesis, $\alpha_j(\lambda_j - \lambda_k) = 0$ for $1 \le j \le k-1$, and since $\lambda_j \neq \lambda_k$, we have $\alpha_j = 0$. Therefore, $\alpha_k v_k = 0$, and since $v_k \neq o$, we have $\alpha_k = 0$. □

Corollary 6. *Let A be an $n \times n$ matrix. If the geometric multiplicity of every eigenvalue of A is equal to its algebraic multiplicity, then A is similar to a diagonal matrix. In particular, this is the case if all eigenvalues of A have algebraic multiplicity 1, i.e., if A has n distinct eigenvalues.*

Example 2. *Let $a_0 = 3$, $a_1 = 8$ and $a_{n+2} = 5a_{n+1} - 6a_n$ for $n \ge 0$. Determine a formula for a_n.*

Let $A = \begin{pmatrix} 0 & 1 \\ -6 & 5 \end{pmatrix}$. Note that $A(a_n, a_{n+1})^T = (a_{n+1}, 5a_{n+1} - 6a_n)^T = (a_{n+1}, a_{n+2})^T$, and thus $(a_n, a_{n+1})^T = A^n(3,8)^T$. The eigenvalues of A are 2 and 3, and thus

$$A = C\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}C^{-1}$$

for some matrix C. Therefore, $(a_n, a_{n+1})^T = C \begin{pmatrix} 2^n & 0 \\ 0 & 3^n \end{pmatrix} C^{-1} \begin{pmatrix} 3 \\ 8 \end{pmatrix}$. It follows that $a_n = \beta_1 2^n + \beta_2 3^n$ for some β_1 and β_2. Since $a_0 = 3$ and $a_1 = 8$, we have $\beta_1 = 1$ and $\beta_2 = 2$. Hence, $a_n = 2^n + 2 \cdot 3^n$.

3 Jordan normal form

Not all matrices are diagonalizable. However, a slight weakening of this claim is true.

Definition 4. *Let $J_k(\lambda)$ be the $k \times k$ matrix $\begin{pmatrix} \lambda & 1 & 0 & 0 & \dots \\ 0 & \lambda & 1 & 0 & \dots \\ & & \dots & & \\ 0 & 0 & \dots & 0 & \lambda \end{pmatrix}$. We call each such matrix a Jordan λ-block.*

A matrix J is in Jordan normal form if

$$J = \begin{pmatrix} J_{k_1}(\lambda_1) & 0 & 0 & \dots \\ 0 & J_{k_2}(\lambda_2) & 0 & \dots \\ & \dots & & \\ 0 & 0 & \dots & J_{k_m}(\lambda_m) \end{pmatrix}$$

for some integers $k_1, \dots, k_m$ and complex numbers $\lambda_1, \dots, \lambda_m$.

Note that $J_1(\lambda) = (\lambda)$, and that $J_k(\lambda)$ has eigenvalue λ with algebraic multiplicity k and geometric multiplicity 1.

Definition 5. *Let $\mathbf{V}$ be a linear space over complex numbers. A chain of generalized eigenvectors for a linear function $f : \mathbf{V} \to \mathbf{V}$ with eigenvalue λ is a sequence of non-zero vectors $v_1, \dots, v_k$ such that $f(v_1) = \lambda v_1$ and $f(v_i) = \lambda v_i + v_{i-1}$ for $i = 2, \dots k$.*

Lemma 7. *Let $\mathbf{V}$ be a linear space over complex numbers of finite dimension n. For every linear function $f : \mathbf{V} \to \mathbf{V}$, there exist chains $C_1, \dots, C_m$ of generalized eigenvectors such that the union of $C_1, \dots, C_m$ is a basis of $\mathbf{V}$.*

Proof. We prove the claim by induction on n. Let λ be an eigenvalue of f, and let $g : \mathbf{V} \to \mathbf{V}$ be defined by $g(x) = f(x) - \lambda x$. Let $\mathbf{W} = \mathrm{Im}(g)$. Since there exists a non-zero eigenvector corresponding to λ, we have $\dim(\mathrm{Ker}(g)) > 0$, and thus $d = \dim(\mathbf{W}) = n - \dim(\mathrm{Ker}(g)) < n$. Note that if $x \in \mathbf{W}$, then $x = g(y)$ for some $y \in \mathbf{V}$, and $f(x) = f(g(y)) = f(f(y) - \lambda y) = f(f(y)) - \lambda f(y) = g(f(y))$, and thus $f(x) \in \mathbf{W}$. Hence, we can consider f as a function from $\mathbf{W}$ to $\mathbf{W}$. By the induction hypothesis, there exist chains $C'_1, \ldots, C'_{m'}$ of generalized eigenvectors of f such that their union $B' = v_1, \ldots, v_d$ is a basis of $\mathbf{W}$. Without loss of generality, the chains C'_1, $\ldots$, C'_q correspond to the eigenvalue λ. Order the elements of the basis B' so that $v_1, \ldots, v_{m'}$ are the last elements of the chains $C'_1, \ldots, C'_{m'}$. For $i = 1, \ldots, q$, let z_i be a vector in $\mathbf{V}$ such that $g(z_i) = v_i$. Let $C_1, \ldots, C_q$ be the chains obtained from $C'_1, \ldots, C'_q$ by adding last elements $z_1, \ldots, z_q$. Let $C_i = C'_i$ for $i = q+1, \ldots, m'$.

Let $x_1, \ldots, x_q$ be the first elements of the chains $C'_1, \ldots, C'_q$. Note that $x_1, \ldots, x_q \in \mathbf{W} \cap \mathrm{Ker}(g)$. Let v be any vector from $\mathbf{W} \cap \mathrm{Ker}(g)$ and let $(\alpha_1, \ldots, \alpha_d)$ be the coordinates of v with respect to B'. Consider any of the chains C corresponding to an eigenvalue μ, and let v_i be its last element such that $\alpha_i \neq 0$. Then the i-th coordinate of $g(v) = f(v) - \lambda v$ is $(\mu - \lambda)\alpha_i$, and since $g(v) = 0$, we conclude that $\mu = \lambda$. Hence, v only has non-zero coordinates in the chains corresponding to the eigenvalue λ. If v_i is in such a chain and it is not its first element, then let v_j be the element of the chain preceding v_j. Then, the j-th coordinate of $g(v)$ is α_i, and thus $\alpha_i = 0$. We conclude that the only coordinates of v that may possibly be non-zero are those corresponding to $x_1, \ldots, x_q$. Therefore, $K = x_1, \ldots, x_q$ forms a basis of $\mathbf{W} \cap \mathrm{Ker}(g)$.

Let $K' = K, u_1, \ldots, u_t$ be a basis of $\mathrm{Ker}(g)$ extending K (where $t = \dim(\mathrm{Ker}(g)) - q = n - d - q$). For $i = m'+1, \ldots, m'+t$, let C_i be the chain consisting of u_i (which is an eigenvector corresponding to λ), and let $m = m' + t$.

We found chains $C_1, \ldots, C_m$ of generalized eigenvectors such that their union contains n vectors. To show that it forms a basis, it suffices to argue that these vectors are linearly independent. Consider any $p = \sum_{i=1}^{q} \alpha_i z_i + \sum_{i=1}^{t} \beta_i u_i + w$ for some $w \in \mathbf{W}$, and let $u = \sum_{i=1}^{t} \beta_i u_i$. Note that $g(p) \in \mathbf{W}$, and observe that for $i = 1, \ldots, q$, the i-th coordinate of $g(p)$ with respect to the basis B' is equal to α_i. Hence, if $p = o$, then $\alpha_1 = \ldots = \alpha_q = 0$, and thus $w = p - u = -u$. Furthermore, $u \in \mathrm{Ker}(g)$, and thus $g(u) = o$, and if $p = o$, then $g(w) = -g(u) = o$, and $w \in \mathrm{Ker}(g) \cap \mathbf{W} = \mathrm{span}(K)$. However, then $\beta_1 = \ldots = \beta_t = 0$ and $w = o$, since K' is a basis of $\mathrm{Ker}(g)$. □

Theorem 8. *Every square matrix A is similar to a matrix in Jordan normal form.*

Proof. Let $f(x) = Ax$. Let $C_1, \ldots, C_m$ be chains of generalized eigenvectors of f forming a basis B of $\mathbf{C}^n$. If $C_1 = v_1, \ldots, v_k$, then $f(v_1) = \lambda v_1$ and $f(v_i) = \lambda v_i + v_{i-1}$ for $i = 2, \ldots, k$ and some eigenvalue λ. Hence, the first column of $[f]_{B,B}$ is λe_1 and the i-th column of $[f]_{B,B}$ is $\lambda e_i + e_{i-1}$ for $i = 2, \ldots, k$. Therefore, the first k columns of $[f]_{B,B}$ are formed by $J_k(\lambda)$ padded from below by zeros. Similarly, we conclude that $[f]_{B,B}$ is in Jordan normal form, with m blocks corresponding to the chains $C_1, \ldots, C_m$. Note that $[f]_{B,B}$ is similar to A. $\square$

Observation 9. *If A is similar to a matrix in Jordan normal form that contains t Jordan λ-blocks of total size m, then λ is an eigenvalue of A with algebraic multiplicity m and geometric multiplicity t. Consequently, the geometric multiplicity of any eigenvalue is at most as large as its algebraic multiplicity.*

Example 3. *Find the Jordan normal form of the matrix*

$$A = \begin{pmatrix} 3 & 1 & 2 \\ 3 & 3 & 4 \\ -2 & -1 & -1 \end{pmatrix}.$$

We know that A has an eigenvalue 1 of algebraic multiplicity 1 and an eigenvalue 2 of algebraic multiplicity 2 and geometric multiplicity 1. Therefore, the Jordan normal form of A is

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

The eigenvectors $(0, -2, 1)$ (eigenvalue 1) and $(1, 1, -1)$ (eigenvalue 2) form the first elements of the chains of generalized eigenvectors. The second element v for the eigenvalue 2 must satisfy $(A - 2I)v = (1, 1, -1)^T$, which has a solution $v = (-1, 0, 1)$. Hence $A = CDC^{-1}$, where

$$C = \begin{pmatrix} 0 & 1 & -1 \\ -2 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix}.$$

Example 4. *Solve the system of linear differential equations*

$$\begin{aligned} f' &= 3f + g + 2h \\ g' &= 3f + 3g + 4h \\ h' &= -2f - g - h \end{aligned}$$

for functions $f, g, h : \mathbf{R} \to \mathbf{R}$.

Note that $\frac{d}{dx}(f, g, h)^T = (f', g', h')^T = A(f, g, h)^T = CDC^{-1}(f, g, h)^T$ for the matrices A, C, and D from Example 3. Equivalently, $\frac{d}{dx}C^{-1}(f, g, h)^T =$

$DC^{-1}(f,g,h)^T$. Let $(f_1,g_1,h_1)^T = C^{-1}(f,g,h)^T$; hence, we need to solve the system $\frac{d}{dx}(f_1,g_1,h_1)^T = D(f_1,g_1,h_1)^T$, i.e.,

$$\begin{aligned} f_1' &= f_1 \\ g_1' &= 2g_1 + h_1 \\ h_1' &= 2h_1 \end{aligned}$$

The general solution for the equation $r' = \alpha r$ is $r(x) = Ce^{\alpha x}$ for any constant C. Hence, $f_1(x) = C_1 e^x$ and $h_1(x) = C_2 e^{2x}$. Then, $g_1' = 2g_1 + C_2 e^{2x}$, which has solution $g_1(x) = C_2 x e^{2x} + C_3 e^{2x}$ for any constant C_3. Therefore, the solution is $(f_1,g_1,h_1) = C_1(e^x,0,0) + C_2(0,xe^{2x},e^{2x}) + C_3(0,e^{2x},0)$, i.e., any element of

$$span((e^x,0,0),(0,xe^{2x},e^{2x}),(0,e^{2x},0)).$$

Hence $(f,g,h)^T = C(f_1,g_1,h_1)^T$ can be any element of

$$span(C(e^x,0,0)^T, C(0,xe^{2x},e^{2x})^T, C(0,e^{2x},0)^T) =$$

$$span\left((0,-2e^x,e^x)^T, ((x-1)e^{2x}, xe^{2x}, (1-x)e^{2x})^T, (e^{2x}, e^{2x}, -e^{2x})^T\right).$$

Observation 10. *For any $n \geq 1$, we have*

$$[J_k(\lambda)]^n = \begin{pmatrix} \lambda^n & \binom{n}{1}\lambda^{n-1} & \binom{n}{2}\lambda^{n-2} & & \dots & \\ 0 & \lambda^n & \binom{n}{1}\lambda^{n-1} & \binom{n}{2}\lambda^{n-2} & & \dots \\ 0 & 0 & \lambda^n & \binom{n}{1}\lambda^{n-1} & \binom{n}{2}\lambda^{n-2} & \dots \\ & & & \dots & & \end{pmatrix}.$$

Definition 6. *For a square matrix A, let*

$$\exp(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

Observation 11. *If $A = CDC^{-1}$, then $\exp(A) = C\exp(D)C^{-1}$,*

$$\exp(J_k(\lambda)) = \begin{pmatrix} e^\lambda & \frac{e^\lambda}{1!} & \frac{e^\lambda}{2!} & & \dots & \\ 0 & e^\lambda & \frac{e^\lambda}{1!} & \frac{e^\lambda}{2!} & & \dots \\ 0 & 0 & e^\lambda & \frac{e^\lambda}{1!} & \frac{e^\lambda}{2!} & \dots \\ & & & \dots & & \end{pmatrix},$$

and

$$\exp(J_k(\lambda)x) = \begin{pmatrix} e^{\lambda x} & \frac{xe^{\lambda x}}{1!} & \frac{x^2e^{\lambda x}}{2!} & & \dots & \\ 0 & e^{\lambda x} & \frac{xe^{\lambda x}}{1!} & \frac{x^2e^{\lambda x}}{2!} & & \dots \\ 0 & 0 & e^{\lambda x} & \frac{xe^{\lambda x}}{1!} & \frac{x^2e^{\lambda x}}{2!} & \dots \\ & & & \dots & & \end{pmatrix},$$

Example 5. *The solutions to a system of differential equations $v' = Av$ are $v(x) \in Col(\exp(Ax))$.*

In Example 4, we have

$$\exp(Ax) = C\exp(Dx)C^{-1} = C\begin{pmatrix} e^x & 0 & 0 \\ 0 & e^{2x} & xe^{2x} \\ 0 & 0 & e^{2x} \end{pmatrix} C^{-1},$$

and thus the set of solutions is

$$(f,g,h)^T \in Col\left[C\begin{pmatrix} e^x & 0 & 0 \\ 0 & e^{2x} & xe^{2x} \\ 0 & 0 & e^{2x} \end{pmatrix}\right].$$

Lemma 12. *For any polynomial p and an $n \times n$ matrix A, if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A listed with their algebraic multiplicities, then $p(\lambda_1), \ldots, p(\lambda_n)$ are the eigenvalues of $p(A)$ listed with their algebraic multiplicities.*

Proof. By Lemma 1, Observation 3 and Theorem 8, it suffices to prove this for matrices in Jordan normal form. Suppose that $A_1, \ldots, A_m$ are the Jordan blocks of A. Then $p(A)$ is a matrix consisting of blocks $p(A_1), \ldots, p(A_m)$ on the diagonal, and the list of eigenvalues of $p(A)$ is equal to the concatenation of the lists of eigenvalues of $p(A_1), \ldots, p(A_m)$. Therefore, it suffices to prove the claim for a Jordan block $J_k(\lambda)$. By Observation 10, the matrix $p(J_k(\lambda))$ is upper triangular and its entries on the diagonal are all equal to $p(\lambda)$, and thus it has eigenvalue $p(\lambda)$ with the algebraic multiplicity k. □

4 Cayley-Hamilton theorem

Theorem 13 (Cayley-Hamilton theorem)**.** *If p is the characteristic polynomial of an $n \times n$ matrix A, then $p(A) = 0$.*

Proof. By Lemma 1, Observation 3 and Theorem 8, it suffices to prove this for matrices in Jordan normal form. Suppose that $A_1, \ldots, A_m$ are the Jordan blocks of A. Then $p(A)$ is a matrix consisting of blocks $p(A_1), \ldots, p(A_m)$ on the diagonal, and p is the product of characteristic polynomials of $A_1, \ldots, A_m$. Hence, it suffices to show that $p_i(A_i) = 0$ for the characteristic polynomial p_i of A_i. However, if $A_i = J_k(\lambda)$, then $p_i(x) = (\lambda - x)^k$, and $p_i(A_i) = (\lambda I - A_i)^k = 0$. □

Example 6. *Let $A = \begin{pmatrix} 3 & 1 & 2 \\ 3 & 3 & 4 \\ -2 & -1 & -1 \end{pmatrix}$. The characteristic polynomial of A is $p(x) = -x^3 + 5x^2 - 8x + 4$.*

Note that $A^2 = AA = \begin{pmatrix} 8 & 4 & 8 \\ 10 & 8 & 14 \\ -7 & -4 & -7 \end{pmatrix}$ *and* $A^3 = A^2A = \begin{pmatrix} 20 & 12 & 24 \\ 26 & 20 & 38 \\ -19 & -12 & -23 \end{pmatrix}$.
We have $p(A) = -A^3 + 5A^2 - 8A + 4I = 0$.

Corollary 14. *Let* A *be an* $n \times n$ *matrix. Then for any* $m \geq 0$, *the matrix* A^m *is a linear combination of* I, A, A^2, $\ldots$, A^{n-1}, *and thus the space of matrices expressible as polynomials in* A *has dimension at most* n. *Furthermore, if* A *is regular, then* A^{-1} *is contained in this space as well.*