BASIC ALGORITHMS IN LINEAR ALGEBRA

STEVEN DALE CUTKOSKY

Matrices and Applications of Gaussian Elimination

1. Systems of Equations. Suppose that A is an $n \times n$ matrix with coefficents in a field F, and $x = (x_1, \ldots, x_n)^T \in F^n$. Let $v \cdot w = v^T w$ be the dot product of the vectors $v, w \in F^n$. Writing $A = (A^1, A^2, \ldots, A^n)$ where $A^i \in F^m$ are the columns of A, we obtain the formula

$$Ax = x_1 A^1 + \cdots + x_n A^n.$$

Writing

$$A = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{pmatrix}$$

where $A_j \in F_n$ are the rows of A, we obtain the formula

$$Ax = \begin{pmatrix} A_1 x \\ A_2 x \\ \vdots \\ A_m x \end{pmatrix} = \begin{pmatrix} A_1^T \cdot x \\ A_2^T \cdot x \\ \vdots \\ A_m^T \cdot x \end{pmatrix}.$$

2. Computation of the inverse of a matrix. Suppose that A is an $n \times n$ matrix. Transform the $n \times 2n$ matrix $(A|I_n)$ into a reduced row echelon form $(C|B)$. A is invertible iff $C = I_n$. If A is invertible, then $B = A^{-1}$.

3. Computation of a basis of the span of a set of row vectors. Suppose that $v_1, \ldots, v_m \in F_n$. Transform the $m \times n$ matrix

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix}$$

into a reduced row echelon form B. The nonzero rows of B form a basis of $\mathrm{Span}(\{v_1, \ldots, v_m\})$.

4. Computation of a subset of a set of column vectors which is a basis of the span of the set. Suppose that $w_1, \ldots, w_n \in F^m$. Transform the $m \times n$ matrix $(w_1, w_2, \ldots, w_n)$ into a reduced row echelon form B. Let $\sigma(1) < \sigma(2) < \ldots < \sigma(r)$ be the indices of the columns B^i of B which contain a leading 1. Then $\{w_{\sigma(1)}, \ldots, w_{\sigma(r)}\}$ is a basis of $\mathrm{Span}(\{w_1, w_2, \ldots, w_n\})$.

5. Extension of a set of linearly independent row vectors to a basis of F_n. Suppose that $w_1, \ldots, w_m \in F_n$ are linearly independent. Let $\{e_1, \ldots, e_n\}$ be the standard basis of F_n. Transform the $m \times n$ matrix

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{pmatrix}$$

into a reduced row echelon form B. Let $\sigma(1) < \sigma(2) < \ldots < \sigma(n-m)$ be the indexes of the columns of B which **do not** contain a leading 1. Then $\{w_1, \ldots, w_m, e_{\sigma(1)}, \ldots, e_{\sigma(n-m)}\}$ is a basis of F_n. (Some different algorithms are given later in the pages on inner product spaces.)

6. Computation of a basis of the solution space of a homogeneous system of equations. Let $A = (a_{ij})$ be an $m \times n$ matrix, and $X = (x_i)$ be a $n \times 1$ matrix of indeterminates. Let $N(A)$ be the null space of the matrix A (the subspace of F^n of all $X \in F^n$ such that $AX = 0_m$). A basis for $N(A)$ can be found by solving the system $AX = 0_m$ using Gaussian elimination to find the general solution, putting the general solution into a column vector and expanding with indeterminate coefficients. The vectors in this expansion are a basis of $N(A)$.

Calculation of the Matrix of a Linear Map

1. Coordinate vectors. Suppose that V is a vector space, with a basis $\beta = \{v_1, \ldots, v_n\}$. Suppose that $v \in V$. Then there is a unique expansion

$$v = c_1 v_1 + \cdots + c_n v_n$$

with $c_i \in \mathbb{R}$. The coordinate vector of v with respect to the basis β is

$$(v)_\beta = (c_1, \ldots, c_n)^T \in M_{n\times 1}.$$

2. The transition matrix between bases. Suppose that V is a vector space, and $\beta = \{v_1, \ldots, v_n\}$, $\beta' = \{w_1, \ldots, w_n\}$ are bases of V. The transition matrix $M_{\beta'}^{\beta}$ from the basis β to the basis β' is the unique $n \times n$ matrix $M_{\beta'}^{\beta}$ which has the property that

$$M_{\beta'}^{\beta}(v)_\beta = (v)_{\beta'}$$

for all $v \in V$. It follows that

$$M_{\beta'}^{\beta} = ((v_1)_{\beta'}, (v_2)_{\beta'}, \ldots, (v_n)_{\beta'}).$$

We have that $M_{\beta}^{\beta'} = (M_{\beta'}^{\beta})^{-1}$, and if β'' is a third basis of V, then $M_{\beta''}^{\beta'} M_{\beta'}^{\beta} = M_{\beta''}^{\beta}$. The $n \times 2n$ matrix $(w_1, w_2, \ldots, w_n, v_1, \ldots, v_n)$ is transformed by elementary row operations into the reduced row echelon form $(I_n, M_{\beta'}^{\beta})$.

3. The matrix of a linear map. Suppose that $F : V \to W$ is a linear map. Let $\beta = \{v_1, \ldots, v_n\}$ be a basis of V, and $\beta' = \{w_1, \ldots, w_m\}$ be a basis of W.

The matrix $M_{\beta'}^{\beta}(F)$ of the linear map F with respect to the bases β of V and β' of W is the unique $m \times n$ matrix $M_{\beta'}^{\beta}(F)$ which has the property that

$$M_{\beta'}^{\beta}(F)(v)_\beta = (F(v))_{\beta'}$$

for all $v \in V$. It follows that

$$M_{\beta'}^{\beta}(F) = ((F(v_1))_{\beta'}, (F(v_2))_{\beta'}, \ldots, (F(v_n))_{\beta'}).$$

If F is the identity map id (so that $V = W$), then $M_{\beta'}^{\beta}(\mathrm{id})$ is the transition matrix $M_{\beta'}^{\beta}$ defined above. Suppose that $G : W \to X$ is a linear map, and β'' is a basis of X. The composition $G \circ F : V \to X$ of F and G can be represented by the diagram

$$V \xrightarrow{F} W \xrightarrow{G} X.$$

We have

$$M_{\beta''}^{\beta}(G \circ F) = M_{\beta''}^{\beta'}(G) M_{\beta'}^{\beta}(F).$$

A particularly important application of this formula is

$$M_{\beta'}^{\beta'}(F) = S^{-1} M_{\beta}^{\beta}(F) S,$$

where $F : V \to V$ is linear, β and β' are bases of V, and $S = M_{\beta}^{\beta'}$.

A convenient method for computing $M_{\beta'}^{\beta}(F)$ is the following. Let β^* be a basis of W which is easy to compute with (such as a standard basis of W). The $m \times (m+n)$ matrix $((w_1)_{\beta^*}, (w_2)_{\beta^*}, \ldots, (w_m)_{\beta^*}, (F(v_1))_{\beta^*}, \ldots, (F(v_n))_{\beta^*})$ is transformed by elementary row operations into the reduced row echelon form $(I_m, M_{\beta'}^{\beta}(F))$.

Inner Product Spaces

1. The Orthogonal Space. Suppose that A is an $m \times n$ matrix with coefficients in a field F. $R(A)$ is the column space of A, and $N(A)$ is the solution space to $Ax = 0$. $R(A)$ is a subspace of F^m, which has a nondegenerate inner product given by the dot product $v \cdot w = v^T w$ for $v, w \in F^m$. For $x \in R^n$, we have the formula

$$\begin{pmatrix} A_1 \\ \vdots \\ A_m \end{pmatrix} x = \begin{pmatrix} A_1^T \cdot x \\ \vdots \\ A_m^T \cdot x \end{pmatrix},$$

and thus we have the formulas

$$N(A) = [R(A^T)]^\perp \text{ and } N(A^T) = R(A)^\perp.$$

2. Pythagoras's Theorem. Suppose that V is a finite dimensional real vector space with positive definite inner product $<,>$. Suppose that $v, w \in V$ and $v \perp w$. Then

$$||v + w||^2 = ||v||^2 + ||w||^2.$$

3. The Gram Schmidt Process. Suppose that V is a finite dimensional real vector space with positive definite inner product $<,>$, and that $\{x_1, \ldots, x_n\}$ is a basis of V. Let

$$\begin{aligned} v_1 &= x_1 \\ u_1 &= \frac{1}{||x_1||} x_1 \\ v_2 &= x_2 - < x_2, u_1 > u_1 \\ u_2 &= \frac{1}{||v_2||} v_2 \\ v_3 &= x_3 - < x_3, u_1 > u_1 - < x_3, u_2 > u_2 \\ u_3 &= \frac{1}{||v_3||} v_3 \\ &\vdots \end{aligned}$$

Then $\{u_1, u_2, \ldots, u_n\}$ is an orthonormal (ON) basis of V.

4. Coordinate Vector With Respect to an ON Basis. Suppose that V is a finite dimensional real vector space with positive definite inner product $<,>$. Suppose that $\beta = \{u_1, \ldots, u_n\}$ is an ON basis of V and $v \in V$. Then

$$(v)_\beta = (< v, v_1 >, < v, v_2 >, \ldots, < v, v_n >)^T.$$

5. Projection Onto a Subspace. Suppose that V is a finite dimensional real vector space with positive definite inner product $<,>$, and that W is a subspace of V. Then $V = W \oplus W^\perp$; that is, every element $v \in V$ has a unique decomposition $V = w + w'$ with $w \in W$ and $w' \in W^\perp$. This allows us to define the projection $\pi_W : V \to W$ by $\pi_W(v) = w$. π_W is a linear map onto W. $\pi_W(v)$ is the element of W which is the closest to v; if $x \in W$ and $x \neq \pi_W(v)$ then

$$||v - \pi_W(v)|| < ||v - x||.$$

$\pi_W(v)$ can be computed as follows. Let $\{u_1, \ldots, u_s\}$ be an orthonormal basis of W. For $v \in V$, let $c_i =< v, u_i >$ be the component of v along u_i. Then

$$\pi_W(v) = \sum_{k=1}^{s} c_k u_k.$$

Now let us restrict to the case where V is $\mathbb{R}^n$ with $< v, w >= v^T w$. Let $\{u_1, \ldots, u_s\}$ be an orthonormal basis of W. Let U be the $n \times s$ matrix $U = (u_1, \ldots, u_s)$. Let $A = UU^T$, an $n \times n$ matrix. The linear map $L_A : \mathbb{R}^n \to \mathbb{R}^n$ is the projection $\pi_W : \mathbb{R}^n \to W$, followed by inclusion of W into $\mathbb{R}^n$.

6. Orthogonal Matrices. An $n \times n$ real matrix A is orthogonal if $A^T A = I_n$. In the following theorem, we view $\mathbb{R}^n$ as an inner product space with the dot product.

Theorem 0.1. *The following are equivalent for an $n \times n$ real matrix A.*

1) *A is an orthogonal matrix.*
2) *$A^T = A^{-1}$.*
3) *The linear map $L_A : \mathbb{R}^n \to \mathbb{R}^n$ preserves length ($||Ax|| = ||x||$ for all $x \in \mathbb{R}^n$).*
4) *The columns of A form an ON basis of $\mathbb{R}^n$*

7. Least Squares Solutions. Let A be an $m \times n$ real matrix and $b \in \mathbb{R}^m$. A least squares solution of the system $Ax = b$ is a vector $x = \hat{x} \in \mathbb{R}^n$ which minimizes $||b - Ax||$ for $x \in \mathbb{R}^n$. The least squares solutions of the system $Ax = b$ are the solutions to the (consistent) system $A^T Ax = A^T b$.

8. Fourier Series. Let $\mathcal{C}[-\pi, \pi]$ be the continuous (real valued) functions on $[-\pi, \pi]$. $\mathcal{C}[-\pi, \pi]$ is a real vector space, with the positive definite inner product

$$< f, g >= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t)g(t)dt.$$

The norm of $f \in \mathcal{C}[-\pi, \pi]$ is defined by $||f||^2 =< f, f >$. For a positive integer n, let T_n be the subspace of $\mathcal{C}[-\pi, \pi]$ which has the orthonormal basis

$$\{\frac{1}{\sqrt{2}}, \sin(t), \cos(t), \sin(2t), \cos(2t), \ldots, \sin(nt), \cos(nt)\}.$$

Suppose that $f \in \mathcal{C}[-\pi, \pi]$. The projection of f on T_n (T is for "trigonometric functions") is

$$f_n(t) = a_0 \frac{1}{\sqrt{2}} + b_1 \sin(t) + c_1 \cos(t) + \cdots + b_n \sin(t) + c_n \cos(nt)$$

where

$$a_0 =< f(t), \frac{1}{\sqrt{2}} >= \frac{1}{\sqrt{2}\pi} \int_{-\pi}^{\pi} f(t)dt$$

and for $1 \leq k \leq n$,

$$b_k =< f(t), \sin(kt) >= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin(kt)dt,$$

$$c_k =< f(t), \cos(kt) >= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos(kt)dt.$$

a_0, b_k, c_k are called the Fourier coefficients of f. f_n is the best approximation of f in T_n, in the sense that $g = f_n$ minimizes $||f - g||$ for $g \in T_n$.

The infinite series

$$g(t) = a_0 \frac{1}{\sqrt{2}} + b_1 \sin(t) + c_1 \cos(t) + \cdots + b_n \sin(t) + c_n \cos(nt) + \cdots$$

converges to f on the interval $[-\pi, \pi]$. $g(t)$ is defined everywhere on $\mathbb{R}$. $g(t)$ is periodic of period 2π; that is $g(a + 2\pi) = g(a)$ for all $a \in \mathbb{R}$. Thus in general, $g(t)$ will only be equal to $f(t)$ on the interval $[-\pi, \pi]$. There is an infinite Parseval's formula,

$$||f||^2 = a_0^2 + b_1^2 + c_1^2 + \cdots .$$

9. Extension of a set of LI vectors to a basis of $\mathbb{R}^n$. Let $v_1, \ldots, v_s$ be a set of linearly independent vectors in $\mathbb{R}^n$. Let

$$A = \begin{pmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_s^T \end{pmatrix}.$$

Let $\{v_{s+1}, \ldots, v_n\}$ be a basis of $N(A)$ (which can be computed using Gaussian elimination). Then $\{v_1, \ldots, v_s, v_{s+1}, \ldots, v_n\}$ is a basis of $\mathbb{R}^n$. Warning: This algorithm does not work in $\mathbb{C}^n$. The reason is that $\mathbb{C}^n$ might not be equal to $W \oplus W^\perp$ if W is a subspace of $\mathbb{C}^n$. To fix this problem, we need the notion of Hermitian inner product. An extension of this algorithm that works over any subfield F of $\mathbb{C}$ will be given below in 11.

10. Hermitian Inner Product Spaces. Suppose that V is a complex vector space. Then the notion of positive definite inner product is generalized to that of Hermitian inner product (warning: an Hermitian inner prouduct is not a bilinear form, so it is not a nondegenerate inner product). The dot product on $\mathbb{C}^n$ is not Hermitian. $\mathbb{C}^n$ has the Hermitian inner product $< v, w >= v^T\overline{w}$ for $v, w \in \mathbb{C}^n$ (here $\overline{w}$ is the complex conjugate of w). The statements of 1 through 9 above all generalize to Hermitian inner products (with $\mathbb{R}$ replaced by $\mathbb{C}$). The statement of 1 for a complex matrix A becomes

$$\begin{pmatrix} A_1 \\ \vdots \\ A_m \end{pmatrix} x = \begin{pmatrix} < A_1^T, \overline{x} > \\ \vdots \\ < A_m^T, \overline{x} > \end{pmatrix} = \begin{pmatrix} < x, \overline{A_1}^T > \\ \vdots \\ < x, \overline{A_m}^T > \end{pmatrix},$$

and thus $R(\overline{A}^T)^\perp = N(A)$ and $R(A)^\perp = N(\overline{A}^T)$.

The projection matrix A of 5 becomes $A = U\overline{U}^T$. The orthogonal matrix defined in 6 generalizes to a unitary matrix. An $n \times n$ complex matrix A is unitary if $\overline{A}^T A = I_n$. The criterion of 2) of the theorem of 6 then becomes $\overline{A}^T = A^{-1}$. The least squares solutions to $Ax = b$ are the solutions to $\overline{A}^T Ax = \overline{A}^T b$. The inner product in 8 becomes

$$< f, g >= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\overline{g(t)}dt.$$

11. Extension of a set of LI vectors to a basis of F^n. Let F be a subfield of $\mathbb{C}$, and let $v_1, \ldots, v_s$ be a set of linearly independent vectors in F^n. Let

$$A = \begin{pmatrix} \overline{v}_1{}^T \\ v_2^T \\ \vdots \\ \overline{v}_s{}^T \end{pmatrix}.$$

Let $\{v_{s+1}, \ldots, v_n\}$ be a basis of $N(A)$ (which can be computed using Gaussian elimination). Then $\{v_1, \ldots, v_s, v_{s+1}, \ldots, v_n\}$ is a basis of F^n.

Eigenvalues and Diagonalization

1. Eigenvalues and Eigenvectors. Suppose that $A \in M_{n\times n}(F)$ is an $n \times n$ matrix. $\lambda \in F$ is an eigenvalue of A if there exists a nonzero vector $v \in F^n$ such that $Av = \lambda v$. Such a v is called an eigenvector of A with eigenvalue λ. For $\lambda \in F$,

$$E(\lambda) = \{v \in F^n \mid Av = \lambda v\}$$

is a subspace of F^n. λ is an eigenvalue of A if and only if $E(\lambda) \neq \{0\}$. The nonzero elements of $E(\lambda)$ are the eigenvectors of A with eigenvalue λ. If λ is an eigenvalue of A, then $E(\lambda)$ is called an eigenspace of A. Thus A is **not** invertible if and only if $\lambda = 0$ is an eigenvalue of A. The eigenspace $E(\lambda)$ is the solution space $N(A - \lambda I_n)$.

2. The Characteristic Polynomial. The characteristic polynomial of $A \in M_{n\times n}(F)$ is $P_A(t) = \mathrm{Det}(tI_n - A)$. Observe that $P_A(t) = (-1)^n \mathrm{Det}(A - tI_n)$. The roots of $P_A(t) = 0$ are the eigenvalues of A.

3. Diagonalization of Matrices. Suppose that $A \in M_{n\times n}(F)$. We say that A is diagonalizable (over F) if A is similar to a diagonal matrix; that is, there exists an invertible $n \times n$ matrix $B \in M_{n\times n}(F)$ such that $B^{-1}AB = D$ is a diagonal matrix.

Let $\beta^* = \{e^1, \ldots, e^n\}$ be the standard basis of F^n. By 2, we have that a matrix $A \in M_{n\times n}(F)$ has only finitely many distinct eigenvalues, say $\lambda_1, \lambda_2, \ldots, \lambda_r$.

Suppose that dim $E(\lambda_i) = s_i$ for $1 \leq i \leq r$. For $1 \leq i \leq r$, let $v_{i,1}, \ldots, v_{i,s_i}$ be a basis of $E(\lambda_i)$. Then

$$v_{1,1}, \ldots, v_{1,s_1}, v_{2,1}, \ldots, v_{r,s_r}$$

is a linearly independent set of vectors (It can be proven that if $w_1, \ldots, w_s$ are eigenvectors for A with distinct eigenvalues, then $w_1 + w_2 + \cdots + w_s = 0$ implies $w_i = 0$ for all i.) If they form a basis β of F^n, then we have an equation

$$M_{\beta}^{\beta^*} A M_{\beta^*}^{\beta} = D = \begin{pmatrix} \lambda_1 & & & & & \\ & \vdots & & & & \\ & & \lambda_1 & & & \\ & & & \lambda_2 & & \\ & & & & \vdots & \\ & & & & & \lambda_r \end{pmatrix}$$

where all nondiagonal entries of D are zero. Thus we have diagonalized A. The matrix

$$M_{\beta^*}^{\beta} = (v_{1,1}, \ldots, v_{1,s_1}, v_{2,1}, \ldots, v_{r,s_r})$$

and $M_{\beta}^{\beta^*} = (M_{\beta^*}^{\beta})^{-1}$.

Working backwards through this construction, we see that an $n \times n$ matrix A is diagonalizable over F if and only if F^n has a basis of eigenvectors of A. In summary, we always have that $s_1 + \cdots + s_r \leq n$, and A is diagonalizable if and only if $s_1 + \cdots + s_r = n$.

4. Eigenvalues and Diagonalization of Operators. Everything above generalizes to an operator $T : V \to V$, where V is an n dimensional vector space over a field F. $\lambda \in F$ is an eigenvalue of T if there exists a nonzero vector $v \in V$ such that $Tv = \lambda v$. Such a v is called an eigenvector of T with eigenvalue λ. We can then form the eigenspace $E(\lambda)$ of an eigenvalue λ of T, which is a subspace of V. Suppose that β is a basis of V. Then we can compute the matrix $M_{\beta}^{\beta}(T)$ of T with respect to the basis β. Further, we can compute the characteristic polynomial $P_{M^\beta_\beta(T)}(t)$ of $M^\beta_\beta(T)$. This polynomial is independent of the choice of basis β of V. Thus we can define the characteristic polynomial of T to be $P_T(t) = P_{M^\beta_\beta(T)}(t)$, computed from any choice of basis β of V. We have that the roots of $P_T(t) = 0$ are the eigenvalues of T.

We say that T is diagonalizable if there exists a basis β of V consisting of eigenvectors of T. In this case, the matrix $M^\beta_\beta(T)$ is a diagonal matrix.

5. Diagonalization of Real Symmetric Matrices. Suppose that $A \in M_{n\times n}(\mathbb{R})$ is a symmetric matrix. Then the spectral theorem tells us that all eigenvalues of A are real and that $\mathbb{R}^n$ has a basis of eigenvectors of A. Further, eigenvectors with distinct eigenvalues are perpendicular. Thus $\mathbb{R}^n$ has an orthonormal basis of eigenvectors. This means that we may refine our diagonalization algorithm of 3, adding an extra step, using Gram Schmidt to obtain an ON basis $u_{i,1}, \ldots, u_{i,s_i}$ of $E(\lambda_i)$ from the basis $v_{i,1}, \ldots, v_{i,s_i}$. Since eigenvectors with distinct eigenvalues are perpendicular, we may put all of these ON sets of vectors together to obtain an ON basis

$$u_{1,1}, \ldots, u_{1,s_1}, u_{2,1}, \ldots, u_{r,s_r}$$

of $\mathbb{R}^n$. Let $U = (u_{1,1}, \ldots, u_{1,s_1}, u_{2,1}, \ldots, u_{r,s_r})$. U is an orthogonal matrix, so that $U^{-1} = U^T$. We have orthogonally diagonalized A,

$$U^T A U = D = \begin{pmatrix} \lambda_1 & & & & & \\ & \vdots & & & & \\ & & \lambda_1 & & & \\ & & & \lambda_2 & & \\ & & & & \vdots & \\ & & & & & \lambda_r \end{pmatrix}$$

where all nondiagonal entries of D are zero, and U is an orthogonal matrix.

6. Triangularization of Matrices. Suppose that $A \in M_{n,n}(F)$. A triangularization of A (over F) is a factorization $P^{-1}AP = T$ where $P, T \in M_{n,n}(F)$, P is invertible and T is upper triangular.

A is triangularizable over F if and only if all of the eigenvalues of A are in F (this will always be true if $F = \mathbb{C}$ is the complex numbers). The following algorithm produces a triangularization of A.

1. Let $v_1, \dots, v_s$ be a maximal set of linearly independent eigenvectors for A, with respective eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_s$. Extend $\{v_1, \dots, v_s\}$ to a basis $v_1, \dots, v_n$ of F^n (This can be done for any F by algorithm 5 in "Matrices and applications", or by algorithm 9 or its extension algorithm 11 in "Inner product spaces" if F is contained in $\mathbb{R}$ or $\mathbb{C}$). Then $P_1 = (v_1, v_2, \dots, v_n)$ satisfies
$$P_1^{-1}AP_1 = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 & * \\ 0 & \lambda_2 & & 0 & * \\ & & \dots & & \\ 0 & 0 & \dots & \lambda_s & * \\ 0 & 0 & \dots & 0 & B \end{pmatrix},$$
where B is an $(n-s) \times (n-s)$ matrix.
2. The eigenvalues of B are a subset of the eigenvalues of A.
3. If $Q^{-1}BQ = S$ is an upper triangular matrix (Q triangularizes B), then
$$P_2 = \begin{pmatrix} I_s & 0_{s\times(n-s)} \\ 0_{(n-s)\times(n-s)} & Q \end{pmatrix}$$
triangularizes A.

Jordan Form

For $\lambda \in \mathbb{C}$, the Jordan block $B_n(\lambda)$ is the $n \times n$ matrix

$$B_n(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & & & & \vdots \\ \vdots & \vdots & & & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}.$$

$B_n(\lambda)$ has the characteristic polynomial

$$P_A(t) = \mathrm{Det}(tI_n - B_n(\lambda)) = (t-\lambda)^n.$$

The only eigenvalue of $B_n(\lambda)$ is λ. The eigenspace of λ for $B_n(\lambda)$ is the solution space to $(B_n(\lambda) - \lambda I_n)X = 0$.

$$B_n(\lambda) - \lambda I_n = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & & & & \vdots \\ \vdots & \vdots & & & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

So the solutions are $x_2 = x_3 = \cdots = x_n = 0$, and a basis of the eigenspace $E(\lambda)$ of $B_n(\lambda)$ consists of the single vector

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}.$$

In particular, $B_n(\lambda)$ is diagonalizable if and only if $n = 1$. In this special case, $B_1(\lambda) = (\lambda)$.

A Jordan Matrix J is a matrix

$$J = \begin{pmatrix} B_{n_{11}}(\lambda_1) & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & B_{n_{1r_1}}(\lambda_1) & 0 & \cdots & 0 \\ 0 & 0 & B_{n_{21}}(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & 0 & \cdots & B_{n_{sr_s}}(\lambda_s) \end{pmatrix}$$

where J is a block (partitioned) matrix whose diagonal elements are the Jordan blocks $B_{n_{ij}}(\lambda_i)$. Set

$$t_i = n_{i1} + n_{i2} + \cdots + n_{ir_i}$$

for $1 \le i \le s$. J is an $n \times n$ matrix where $n = t_1 + t_2 + \cdots + t_s$. The characteristic polynomial of J is

$$P_J(t) = \mathrm{Det}(tI_n - J) = (t-\lambda_1)^{t_1}(t-\lambda_2)^{t_2}\cdots(t-\lambda_s)^{t_s}.$$

The eigenvalues of J are $\lambda_1, \cdots, \lambda_s$. Let $e(i)$ be the column vector of length n with a 1 in the ith place and zeros everywhere else.

A basis for $E(\lambda_1)$ is

$$\{e(1), e(n_{11}+1), \ldots, e(n_{11}+\cdots+n_{1,r_1-1}+1)\}.$$

A basis for $E(\lambda_2)$ is

$$\{e(t_1+1), \ldots, e(t_1+n_{21}+\cdots+n_{2,r_2-1}+1)\}$$

and a basis of $E(\lambda_s)$ is

$$\{e(t_1+\cdots+t_{s-1}+1), \ldots, e(t_1+\cdots+t_{s-1}+n_{s1}+\cdots+n_{s,r_s-1}+1)\}.$$

In particular, $E(\lambda_i)$ has dimension r_i, the number of Jordan blocks of J with eigenvalue λ_i.

Example 1.

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

A is a Jordan matrix with 3 Jordan blocks:

$$\begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}, \ (2), \ \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

Theorem 0.2. *Every square matrix A with complex coefficients is similar to a Jordan Matrix J; that is, there is an invertible complex matrix C such that*

$$J = C^{-1}AC.$$

J is called a Jordan form of A.

The Jordan form of a matrix A is uniquely determined, up to permuting the Jordan blocks of a Jordan form.

This theorem fails over the reals. Even if A is a real matrix, it will in general not be similar to a real Jordan matrix. The essential point that makes everything work out over the complex numbers is the "fundamental theorem of algebra" which states that a nonconstant polynomial with complex coefficients has a complex root, so that it must factor into a product of linear factors (with complex coefficients). Thus every complex matrix has a complex eigenvalue (since the characteristic polynomial must have a complex root). However, there are real matrices which do not have a real eigenvalue.

Example 2. Suppose that $P_A(t) = (t-2)^2(t+3)^2$. Then A has (up to permuting Jordan blocks) one of the following Jordan forms:

$$F_1 = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix}, \ F_2 = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix},$$

$$F_3 = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -3 & 1 \\ 0 & 0 & 0 & -3 \end{pmatrix}, \quad F_4 = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -3 & 1 \\ 0 & 0 & 0 & -3 \end{pmatrix}.$$

Suppose that A is an $n \times n$ matrix with complex coefficients. Let J be a Jordan form of A (with all of the above notation), so that $P_A(t) = P_J(t)$. There is a factorization

$$P_A(t) = (t - \lambda_1)^{t_1}(t - \lambda_2)^{t_2} \cdots (t - \lambda_s)^{t_s}$$

where λ_i are the distinct complex eigenvalues of A, and $t_1 + t_2 + \cdots + t_s = n$. The algebraic multiplicity of A for λ_i is t_i, and the geometric multiplicity of A for λ_i is dim $E(\lambda_i)$, the dimension of the eigenspace of λ_i for A. For each eigenvalue λ_i of A, we have

$$1 \le \dim E(\lambda_i) \le t_i.$$

A is diagonalizable if and only if we have equality of the algebraic and geometric multiplicities for all eigenvalues λ_i of A.

A polynomial $f(t) \in \mathbb{C}[t]$ is monic if its leading coefficient is 1; that is, $f(t)$ has the form $f(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_0$ with $a_0, a_1, \ldots, a_{n-1} \in \mathbb{C}$. The minimal polynomial $q_A(t)$ of A is the (unique) monic polynomial in $\mathbb{C}[t]$ which has the property that $q_A(A) = 0$, and if $f(t) \in \mathbb{C}[t]$ satisfies $f(A) = 0$, then $q_A(t)$ divides $f(t)$. If B is similar to A, then $q_B(t) = q_A(t)$, so that $q_A(t) = q_J(t)$. Let $\varphi(i) = \max\{n_{ij} \mid 1 \le i \le r_i\}$. Then

$$q_A(t) = (t - \lambda_1)^{\varphi(1)}(t - \lambda_2)^{\varphi(2)} \cdots (t - \lambda_s)^{\varphi(s)}.$$

The Cayley-Hamilton theorem tells us that $p_A(A) = 0$, so that $q_A(t)$ divides $p_A(t)$. This gives us a method of computing $q_A(t)$. Assuming that we are able to factor the characteristic polynomial of a matrix A, we can thus calculate fairly easily a lot of information about the Jordan form. For matrices of small size, just knowing the characteristic polynomial, the minimal polynomial, and the geometric multiplicities will often uniquely determine the Jordan form. Of course, this is not enough information to compute the Jordan form for general matrices!

Exercises on Jordan Form.

1. Which of the following are Jordan matrices?

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}, \quad \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 2 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

2. What are the possible Jordan forms of A (up to permutation of Jordan blocks) if A has the given characteristic polynomial?
   a) $P_A(t) = (t-4)^2 t(t+2)^2$.
   b) $P_A(t) = (t-2)^3$.
3. Suppose that A is a 4×4 matrix with eigenvalues 2 and 5. Suppose that $E(2)$ has dimension 1 and $E(5)$ has dimension 3. What are the possible Jordan forms of A (up to permutation of Jordan blocks)?