MODULE 11

Topics: Hermitian and symmetric matrices

Setting: A is an $n \times n$ real or complex matrix defined on $\mathbb{C}_n$ with the

$$\text{complex dot product } \langle x, y\rangle = \sum_{j=1}^{n} x_j \overline{y_j}.$$

Notation: $A^* = \overline{A^T}$, i.e., $a_{ij} = \overline{a_{ji}}$.

We know from Module 4 that

$$\langle Ax, y\rangle = \langle x, A^* y\rangle$$

for all $x, y \in \mathbb{C}_n$.

Definition: If $A = A^T$ then A is symmetric.

If $A = A^*$ then A is Hermitian.

Examples: $\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$ is symmetric but not Hermitian

$\begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}$ is Hermitian but not symmetric.

Note that a real symmetric matrix is Hermitian.

Theorem: *If A is Hermitian then*

$$\langle Ax, x\rangle \text{ is real for all } x \in \mathbb{C}_n.$$

Proof: $\langle Ax, x\rangle = \overline{\langle x, Ax\rangle} = \overline{\langle Ax, x\rangle}$, so the number $\langle Ax, x\rangle$ is real.

Theorem: *If A is Hermitian then its eigenvalues are real.*

Proof: Suppose that $Au = \lambda u$, then

$$\lambda\langle u, u\rangle = \langle \lambda u, u\rangle = \langle Au, u\rangle = \langle u, A^* u\rangle = \langle u, Au\rangle = \langle u, \lambda u\rangle = \bar{\lambda}\langle u, u\rangle.$$

It follows that $\lambda = \bar{\lambda}$ so that λ is real.

Theorem: *If A is Hermitian then eigenvectors corresponding to distinct eigenvalues are orthogonal.*

Proof: Suppose that $Au = \lambda u$ and $Av = \mu v$ for $\lambda \neq \mu$. Then

$$\lambda\langle u, v\rangle = \langle Au, v\rangle = \langle u, Av\rangle = \langle u, \mu v\rangle = \mu\langle u, v\rangle.$$

Since $\mu \neq \lambda$ it follows that $\langle u, v \rangle = 0$.

Suppose that A has n distinct eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ with corresponding orthogonal eigenvectors $\{u_1, \ldots, u_n\}$. Let us also agree to scale the eigenvectors so that

$$\langle u_i, u_j \rangle = \delta_{ij}$$

where δ_{ij} is the so-called Kronecker delta

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$$

We recall that the eigenvalue equation can be written in matrix form

$$AU = U\Lambda$$

where $U = (u_1 \ u_2 \ \ldots \ u_n)$ and $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$. We now observe from the orthonormality of the eigenvectors that

$$U^* U = I.$$

Hence $U^{-1} = U^*$ and consequently

$$A = U \Lambda U^*.$$

In other words, A can be diagonalized in a particularly simple way.

Example: Suppose

$$A = \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}$$

We saw that A is Hermitian; its eigenvalues are the roots of $(1-\lambda)^2 - 1 = 0$ so that

$$\lambda_1 = 0, \qquad u_1 = (1, i)/\sqrt{2}$$

$$\lambda_2 = 2, \qquad u_2 = (1, -i)/\sqrt{2}$$

which shows that $\langle u_1, u_2 \rangle = 0$. Thus

$$U^* U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$A = U \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} U^*.$$

We saw from the example

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

that the eigenvector system for A can only be written in the form

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

This matrix A cannot be diagonalized because we do not have n linearly independent eigenvectors. However, a Hermitian matrix can always be diagonalized because we can find an orthonormal eigenvector basis of $\mathbb{C}_n$ regardless of whether the eigenvalues are distinct or not.

Theorem: *If A is Hermitian then A has n orthonormal eigenvectors $\{u_1, \dots, u_n\}$ and*

$$A = U \Lambda U^*$$

where

$$\Lambda = \mathrm{diag}\{\lambda_1, \dots, \lambda_n\}.$$

Proof: If all the eigenvalues were distinct then the results follows from the forgoing discussion. Here we shall outline a proof this result for the case where A is real so that the setting is $\mathbb{R}_n$ rather than $\mathbb{C}_n$. Define the function

$$f(y) = \langle Ay, y \rangle$$

Analysis tells us that this function has a minimum on the set $\langle y, y \rangle = 1$. The minimizer can be found with Lagrange multipliers. We have the Lagrangian

$$\mathcal{L}(y) = \langle Ay, y \rangle - \lambda(\langle y, y \rangle - 1)$$

and the minimizer x_1 is found from the necessary condition

$$\nabla \mathcal{L}(y) = 0.$$

We compute

$$\frac{\partial \mathcal{L}}{\partial y_k} = \langle A_k, y \rangle + \langle Ay, e_k \rangle - 2\lambda y_k = 2(\langle A_k, y \rangle - \lambda y_k)$$

so that we have to solve

$$\nabla \mathcal{L} \equiv Ay - \lambda y = 0$$

$$\langle y, y \rangle = 1.$$

The solution to this problem is the eigenvector x_1 with eigenvalue λ_1.

If we now apply the Gram-Schmidt method to the set of $n+1$ dependent vectors $\{x_1, e_1, \ldots, e_n\}$ we end up with n orthogonal vectors $\{x_1, y_2, \ldots, y_n\}$. We now minimmize

$$f(y) = \langle Ay, y \rangle \text{ subject to } \|y\| = 1$$

over all $y \in \text{span}\{y_2, \ldots, y_2\}$ and get the next eigenvalue λ_2 and eigenvector x_2 which then necessarily satisfies $\langle x_1, x_2 \rangle = 0$. Next we find an orthogonal basis of $\mathbb{R}_n$ of the form $\{x_1, x_2, y_3, \ldots, y_n\}$ and minimize $f(y)$ over span $\{y_3, \ldots, y_n\}$ subject to the constraint that $\|y\| = 1$. We keep going and find eventually n eigenvectors, all of which are mutually orthogonal. Afterwards all eigenvectors can be normalized so that we have an orthonormal eigenvector basis of $\mathbb{R}_n$ which we denote by $\{u_1, \ldots, u_n\}$.

If A is real and symmetric then this choice of eigenvectors satisfies

$$U^T U = I.$$

If A is complex then the orthonormal eigenvectors satisfy

$$U^* U = I.$$

In either case the eigenvalue equation

$$AU = U\Lambda$$

can be rewritten as

$$A = U \Lambda U^* \text{ or equivalently, } U^* A U = \Lambda.$$

We say that A can be diagonalized.

We shall have occasion to use this result when we talk about solving first order systems of linear ordinary differential equations.

The diagonalization theorem guarantees that for each eigenvalue one can find an eigenvector which is orthogonal to all the other eigenvectors regardless of whether the eigenvalue is distinct or not. This means that if an eigenvalue μ is a repeated root of multiplicity k (meaning that $\det(A - \lambda I) = (\lambda - \mu)^k g(\lambda)$, where $g(\mu) \neq 0$), then $\dim \mathcal{N}(A - \mu I) = k$. We simply find an orthogonal basis of this null space. The process is, as usual, the application of Gaussian elimination to

$$(A - \mu I)x = 0$$

and finding k linearly independent vectors in the null space which can then be made orthogonal with the Gram-Schmidt process.

Some consequences of this theorem: Throughout we assume that A is Hermitian and that x is not the zero vector.

Theorem: $\langle Ax, x\rangle > 0$ *if and only if all eigenvalues of A are strictly positive.*

Proof: Suppose that $\langle Ax, x\rangle > 0$. If λ is a non-positive eigenvalue with eigenvector u then $\langle Au, u\rangle = \lambda\langle u, u\rangle \leq 0$ contrary to assumption. Hence the eigenvalues must be positive. Conversely, assume that all eigenvalues are positive. Let x be arbitrary, then the existence of an orthonormal eigenvector basis $\{u_j\}$ assures that

$$x = \alpha_1 u_1 + \cdots + \alpha_n u_n.$$

If we substitute this expansion into $\langle Ax, x\rangle$ we obtain

$$\langle Ax, x\rangle \sum_{j=1}^{n} \lambda_j \alpha_j \overline{\alpha_j} > 0$$

for any non-zero x.

Definition: A is positive definite if $\langle Ax, x\rangle > 0$ for $x \neq 0$.

A is positive semi-definite if $\langle Ax, x\rangle \geq 0$ for all x.

A is negative definite if $\langle Ax, x\rangle < 0$ for $x \neq 0$.

A is negative semi-definite if $\langle Ax, x\rangle \leq 0$ for all x.

Note that for a positive or negative definite matrix $Ax = 0$ if and only if $x = 0$ while a semi-definite matrix may have a zero eigenvalue.

An application: Consider the real quadratic polynomial

$$P(x,y,z) = a_{11}x^2 + a_{12}xy + a_{13}xz + a_{22}y^2 + a_{23}yz + a_{33}z^2 + b_1x + b_2y + b_3z + c$$

We "know" from analytic geometry that $P(x,y,z) = 0$ describes an ellipsoid, paraboloid or hyperboloid in $\mathbb{R}_3$. But how do we find out? We observe that $P(x,y,z) = 0$ can be rewritten as

$$P(x,y,z) = \left\langle A \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right\rangle + \left\langle b, \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right\rangle c = 0$$

where

$$A = \begin{pmatrix} a_{11} & a_{12}/2 & a_{13}/2 \\ a_{12}/2 & a_{22} & a_{23}/2 \\ a_{13}/2 & a_{23}/2 & a_{33} \end{pmatrix} \qquad \text{is a symmetric real matrix.}$$

Hence A can be diagonalized

$$A = U\Lambda U^T$$

Define the new vector (new coordinates)

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = U^T \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

Then

$$\begin{aligned} P(X,Y,X) &= \left\langle \Lambda \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}, \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \right\rangle + \left\langle b, U \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \right\rangle + c \\ &= \lambda_1 X^2 + \lambda_2 Y^2 + \lambda_3 Z^2 + \text{lower order terms} \end{aligned}$$

where $\{\lambda_j\}$ are the eigenvalues of A. If all are positive or negative we have an ellipsoid, if all are non-zero but not of the same sign we have a hyperboloid, and if at least one eigenvalue is zero we have a paraboloid. The lower order terms determine where these shapes are centered but not their type.

Module 11 Homework

1) Suppose that A has the property $A = -A^*$. In this case A is said to be skew-Hermitian.
   i) Show that all eigenvalues of A have to be purely imaginary.
   ii) Prove or disprove: The eigenvectors corresponding to distinct eigenvalues of a skew-Hermitian matrix are orthogonal with respect to the complex dot product.
   iii) If A is real and skew-Hermitian what does this imply about A^T?
2) Let A be a skew-Hermitian matrix.
   i) Show that for each $x \in \mathbb{C}_n$ we have $Re\langle Ax, x\rangle = 0$
   ii) Show that for any matrix A we have

$$\langle Ax, x\rangle = \langle Bx, x\rangle + \langle Cx, x\rangle$$

   where B is Hermitian and C is skew-Hermitian.
3) Suppose that $U^*U = I$. What can you say about the eigenvalues of U?
4) Find a new coordinate system such that the conic section

$$x^2 - 2xy + 4y^2 = 6$$

   is in standard form so that you can read off whether it is an ellipse, parabola or hyperbola.