MATH S-15, SUMMER 2001
GROUPS, GRAPHS, AND ALGEBRAIC STRUCTURES FOR COMPUTING
Lecture # 6, supplement

Invertible Matrices

Just as any one-to-one function has an inverse, an $n \times n$ matrix A will be invertible if, when considered as a linear transformation, it has an inverse. That is, we would need to be able to solve the vector equation $A\mathbf{v} = \mathbf{b}$ for any given vector $\mathbf{b} \in \mathbb{R}^n$. When this is the case, the inverse transformation is also linear and may be represented by another $n \times n$ matrix, which we denote A^{-1}.

To begin, we consider the simplest case, that of the 2×2 matrix, and we let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. If $\mathbf{b} = \begin{bmatrix} z \\ w \end{bmatrix}$, then we need to be able to solve the equation $A\mathbf{v} = \mathbf{b}$ for the general vector $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}$. This yields the system of two equations in two unknowns (recall that a, b, c, d, z, and w are given):

$$ax + by = z$$

$$cx + dy = w$$

Solving for x and y yields:

$$x = \frac{dz - bw}{ad - bc} \quad \text{and} \quad y = \frac{-cz + aw}{ad - bc}$$

Another way to write this in terms of the vectors $\mathbf{v}$ and $\mathbf{b}$ is:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} dz - bw \\ -cz + aw \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} z \\ w \end{bmatrix}$$

or as $\mathbf{v} = A^{-1}\mathbf{b}$, where

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Thus, we have constructed by hand the inverse for the matrix A, *so long as* $ad - bc \neq 0$. In the case that $ad - bc = 0$, the simultaneous equations above have either infinitely many solutions or no solutions at all, either of which implies that the linear transformation corresponding to A does not have an inverse.

Of course, the expression $ad - bc$ is familiar as the determinant of the matrix A, and this is no accident. In fact, this is a small example of the much larger principle:

Theorem: If A is an $n \times n$ matrix, then A is invertible if and only if $\det(A) \neq 0$.

Thus, in the 2×2 case, we write $A^{-1} = \dfrac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. The formula for A^{-1} is much more complicated in the $n \times n$ case, and the reader is referred to the textbook, but the scalar $(\det A)^{-1}$ is still present.

We list a few facts about the inverses and transposes of square matrices. Let A and B be invertible $n \times n$ matrices. Then:

- Fact: $(AB)^{-1} = B^{-1}A^{-1}$
- Fact: $(A^t)^{-1} = (A^{-1})^t$
- Fact: $\det(A^{-1}) = \frac{1}{\det(A)}$
- Fact: $\det(A) = \det(A^t)$

The first of these is a general result from group theory when we have non-commutative groups, and the third of these follows directly from the facts that $AA^{-1} = I$ and that $\det(AB) = \det(A)\det(B)$.

The General Linear Group and its Subgroups

We now move to consider the collection of all invertible $n \times n$ matrices as a group. We define the **general linear group** to be:

$$GL(n, \mathbb{R}) = \{A \in M(n, \mathbb{R}) | \det(A) \neq 0\}$$

This is a group whose operation is matrix multiplication. Closure follows from the fact that $\det(AB) = \det(A)\det(B)$. The identity is the matrix with 1's on the main diagonal and 0's elsewhere. Thus, in the 2×2 case, the identity is: $I = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$. Inverses exist and are computed as we have just done in the previous section.

The collection of matrices that preserve the volume (and orientation) of geometric objects in $\mathbb{R}^n$ form a subgroup of the general linear group known as the **special linear group**:

$$SL(n, \mathbb{R}) = \{A \in GL(n, \mathbb{R}) | \det(A) = 1\}$$

The collection of matrices that preserve the lengths of vectors in $\mathbb{R}^n$ form a subgroup of the general linear group known as the **orthogonal group**:

$$O(n, \mathbb{R}) = \{A \in GL(n, \mathbb{R}) | \; ||A\mathbf{v}|| = ||\mathbf{v}|| \text{ for all } \mathbf{v} \in \mathbb{R}^n\}$$

We list some facts concerning these orthogonal matrices, and we abbreviate the group's name as $O(n)$:

- Fact: If $A \in O(n)$, then $A\mathbf{u} \cdot A\mathbf{v} = \mathbf{u} \cdot \mathbf{v}$.
- Fact: If $A \in O(n)$, then $A^{-1} \in O(n)$.
- Fact: If $A \in O(n)$, then $A^{-1} = A^t$.
- Fact: If $A \in O(n)$, then $\det(A) = \pm 1$.

The first of these says that angles are also preserved by orthogonal transformations, and we note that the third of these is extremely useful as the computation of inverses is generally difficult, whereas the computation of transposes is very simple.

> Examples: Several examples of orthogonal matrices are reflections such as $A = \left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right]$, rotations such as $B = \left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$, and permutations such as $C = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$.

The Eigenvalues and Eigenvectors of a Square Matrix

For all that follows, let $A \in M_n(\mathbb{R})$ be a square $n \times n$ matrix with real entries. Then, as we have seen, A corresponds to a linear transformation $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$.

Linear transformations take vectors from one vector space to another, but a linear transformation as given above can be thought to move vectors around in a single vector space. In this case, some of the types of geometric behavior that we have observed include rotations, reflections, shears, and scaling. The easiest of these to understand is scaling (stretching or shrinking), and we will describe these in much greater detail.

Let A be as above, and let $\mathbf{v}$ be a vector in $\mathbb{R}^n$. We say that $\mathbf{v}$ is an **eigenvector** for A if there is some non-zero scalar λ such that $A\mathbf{v} = \lambda\mathbf{v}$, and in this case, the scalar λ is known as an **eigenvalue** for A. Note that this definition is identifying vectors on which A acts by scalar multiplication, which is much simpler in general than matrix mutliplication.

> Example: Let $A = \left[\begin{smallmatrix} 3 & 4 \\ 3 & 2 \end{smallmatrix}\right]$. Then $\mathbf{v} = \left[\begin{smallmatrix} 4 \\ 3 \end{smallmatrix}\right]$ is an eigenvector for A with eigenvalue 6 because $A\mathbf{v} = \left[\begin{smallmatrix} 24 \\ 18 \end{smallmatrix}\right]$, which equals $6\mathbf{v}$.

One of the first observations that one makes is that if $\mathbf{v}$ is an eigenvector for the matrix A with eigenvalue λ, then so is $c\mathbf{v}$ for any scalar c. In fact, it is easy to show that the set of eigenvectors with a common eigenvalue form a vector space.

In the example above, it was easy to verify that $\mathbf{v}$ was an eigenvector for A, but in order for these concepts to be more useful, we will need a procedure for *finding* eigenvalues and eigenvectors, so we analyze a matrix as follows. Suppose that a non-zero vector $\mathbf{v}$ is an eigenvector for A with eigenvalue λ. Then, if I is the $n \times n$ identity matrix, we have:

$$\begin{aligned} A\mathbf{v} &= \lambda\mathbf{v} \\ A\mathbf{v} &= (\lambda I)\mathbf{v} \\ (A - \lambda I)\mathbf{v} &= \mathbf{0} \end{aligned}$$

But if $\mathbf{v} \neq \mathbf{0}$, then the matrix $A - \lambda I$ must not be invertible. Our condition for the invertibility of a matrix is that its determinant must be non-zero. Thus, the condition for λ to be an eigenvalue for A is that:

$$\det(A - \lambda I) = 0.$$

This condition is important enough to merit another definition, and we define the **characteristic polynomial** of A to be the polynomial (with variable λ) $f_A(\lambda) = \det(A - \lambda I)$. To find the eigenvectors corresponding to the eigenvalue λ thus found, we must solve the system of linear equations given by the vector equation $(A - \lambda I)\mathbf{v} = \mathbf{0}$.

Example: We return to the example above, where $A = \left[\begin{smallmatrix} 3 & 4 \\ 3 & 2 \end{smallmatrix}\right]$.

Then $A - \lambda I = \left[\begin{smallmatrix} 3-\lambda & 4 \\ 3 & 2-\lambda \end{smallmatrix}\right]$, and

$$\begin{aligned} f_A(\lambda) &= \det(A - \lambda I) \\ &= (3-\lambda)(2-\lambda) - 4 \cdot 3 \\ &= \lambda^2 - 5\lambda - 6 \\ &= (\lambda - 6)(\lambda + 1). \end{aligned}$$

Setting $f_A(\lambda) = 0$, we find solutions $\lambda_1 = 6$ and $\lambda_2 = -1$. (Note that the procedure has found a previously undiscovered eigenvalue!) To find the eigenvectors corresponding to the eigenvalue $\lambda_1 = 6$, we continue as follows. We know that $A - 6I$ is not invertible and that an eigenvector must solve the equation $(A - 6I)\mathbf{v} = \mathbf{0}$. Note that $A - 6I = \left[\begin{smallmatrix} -3 & 4 \\ 3 & -4 \end{smallmatrix}\right]$. Let $\mathbf{v} = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ be an arbitrary vector satisfying $(A - 6I)\mathbf{v} = \mathbf{0}$. This yields the pair of simultaneous equations:

$$-3x + 4y = 0 \text{ and } 3x - 4y = 0$$

We discover that this system does not have a unique solution! In fact, for any x and y with the relationship that $x = \frac{4}{3}y$, we get a vector $\mathbf{v}$. But this is what we should expect, since we already know that any scalar multiple of an eigenvector is also an eigenvector. Thus our *set* of eigenvectors corresponding to the eigenvalue $\lambda = 6$ is the set of all vectors of the form $\mathbf{v} = \left[\begin{smallmatrix} \frac{4}{3}y \\ y \end{smallmatrix}\right]$, which is equal to the set of vectors $\left\{ \left[\begin{smallmatrix} \frac{4}{3}y \\ y \end{smallmatrix}\right] \middle| \, y \in \mathbb{R} \right\}$, which is equal to the set of all scalar multiples of the vector $\mathbf{v}_1 = \left[\begin{smallmatrix} 4 \\ 3 \end{smallmatrix}\right]$.

As for the eigenvalue $\lambda_2 = -1$, a similar computation with the matrix $A + I = \left[\begin{smallmatrix} 4 & 4 \\ 2 & 2 \end{smallmatrix}\right]$ yields the pair of simultaneous equations

$$4x + 4y = 0 \text{ and } 2x + 2y = 0$$

which has solution set $\{ \left[\begin{smallmatrix} -y \\ y \end{smallmatrix}\right] | \, y \in \mathbb{R} \}$, which is equal to the set of all scalar multiples of the vector $\mathbf{v}_2 = \left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right]$.

We generalize this procedure in the 2×2 case. When $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$, we get the matrix $A - \lambda I = \left[\begin{smallmatrix} a-\lambda & b \\ c & d-\lambda \end{smallmatrix}\right]$, the characteristic polynomial becomes

$$\begin{aligned} f_A(\lambda) &= (a-\lambda)(d-\lambda) - b \cdot c \\ &= \lambda^2 - (a+d)\lambda + (ad - bc) \\ &= \lambda^2 - (\mathrm{tr}A)\lambda + (\det A) \end{aligned}$$

Note that as a quadratic polynomial over the real numbers, we can expect three cases: two distinct real roots, one doubled real root, or two complex roots which are complex conjugates of each other. Thus it is quite possible for a matrix to have no eigenvalues in the field over which it takes its entries!

Finding eigenvectors is not significantly simpler in general than it was in the example, and we do not provide a procedure.

A few comments about the case when A is an $n \times n$ matrix are in order. Most of the procedure outlined above carries over: the eigenvalues and eigenvectors and the characteristic polynomial are all defined as before, but of course, the computation of all of these becomes more complicated. One notion which does not carry over quite so beautifully is the simple form of the characteristic polynomial in terms of the trace and determinant, though we do get a similar result:

$$f_A(\lambda) = \lambda^n - (\mathrm{tr}A)\lambda^{n-1} + \cdots + (-1)^n(\det A)$$

The characteristic polynomial still has degree n, and the trace and determinant are coefficients of it, but they do not provide the entire formula. Of course, it should also be said that the roots of high degree polynomials are not, in general, easy to find, and when $n > 4$, there does not exist a finite algorithm to do so!