Supplementary Material for Math 3200-105 (13894) Applied Linear Algebra Fall 2016

Martin J. Mohlenkamp

November 29, 2016

This document contains material needed for Math 3200 Applied Linear Algebra but not in the textbook **Matrix Methods: Applied linear Algebra**, third edition, by Richard Bronson and Gabriel Costa, Academic Press 2009. Other uses are discouraged. The material is found in other textbooks (and Wikipedia) and is not original.

Contents

1 Notes on Section 6.5 Linearly Independent Eigenvectors

There is some terminology missing from our book that is used in the WeBWork homework for section 6.5. We will also need it for our supplementary material.

Definition 1. *The* **characteristic polynomial** *of* $\mathbf{A}$ *is* $p(\lambda) = |\mathbf{A} - \lambda\mathbf{I}|$.

Section 6.2 introduces the characteristic equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$ but dose not name the polynomial.

Definition 2. *The* **algebraic multiplicity** *of an eigenvalue* λ_i *is the multiplicity of* λ_i *as a root of the characteristic polynomial.*

In other words, if we factor $p(\lambda)$ and get a factor $(\lambda - \lambda_i)^k$, then the algebraic multiplicity is k. In section 6.2 the algebraic multiplicity is just called multiplicity.

Definition 3. *The* **eigenspace** *corresponding to an eigenvalue* λ *of a matrix* $\mathbf{A}$ *is the set of all vectors* $\mathbf{x}$ *such that* $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$.

Definition 4. *The* **dimension** *of a space is the maximal number of linearly independent vectors that can be selected from it.*

In section 6.5 this number is counted, but was not given a name.

Definition 5. *A* **basis** *of an eigenspace is a linearly independent spanning set for the eigenspace.*

We saw basis in section 2.5 problem 32. The number of vectors in a basis must equal the dimension of the space. If the dimension of an eigenspace is k and you find k linearly independent vectors, these vectors will be a basis.

Definition 6. *The* **geometric multiplicity** *of an eigenvalue λ_i is the dimension of the eigenspace corresponding to λ_i.*

2 Diagonalization

Definition 7. *A matrix* $\mathbf{A}$ *and a matrix* $\mathbf{B}$ *are* **similar** *if there exists an invertible matrix* $\mathbf{P}$ *such that* $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$.

Note that $\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1} = \mathbf{Q}^{-1}\mathbf{B}\mathbf{Q}$ so similarity is an equivalence relation.

Lemma 8. *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are similar then they have the same eigenvalues.*

Proof. Suppose $\mathbf{A}$ has an eigenvalue λ with eigenvector $\mathbf{x}$, so we know $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. Then

$$\mathbf{B}(\mathbf{P}^{-1}\mathbf{x}) = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}\mathbf{P}^{-1}\mathbf{x} = \mathbf{P}^{-1}\mathbf{A}\mathbf{x} = \mathbf{P}^{-1}\lambda\mathbf{x} = \lambda(\mathbf{P}^{-1}\mathbf{x})\,, \tag{1}$$

so λ is an eigenvalue of $\mathbf{B}$ with eigenvector $\mathbf{P}^{-1}\mathbf{x}$. □

Theorem 9. *Similar matrices have the same*

- *rank;*
- *characteristic polynomial and thus the same*
  - *eigenvalues with the same algebraic multiplicities,*
  - *determinant, and*
  - *trace; and*
- *geometric multiplicities of eigenvalues.*

We will not try to prove this theorem. It motivates transforming $\mathbf{A}$ to a similar matrix that is easier to work with and understand. The simplest such matrix is a diagonal matrix.

Definition 10. *A matrix* $\mathbf{A}$ *is* **diagonalizable** *if it is similar to a diagonal matrix.*

Unfortunately, not all matrices are diagonalizable.

Theorem 11. *An $n \times n$ matrix* $\mathbf{A}$ *is* **diagonalizable** *if and only if it has n linearly independent eigenvectors.*

Proof. Suppose we have linearly independent eigenvectors $\{\mathbf{x}_i\}_{i=1}^n$. Combine them as columns into a matrix $\mathbf{P}$. Since the columns of $\mathbf{P}$ are linearly independent, $\text{rank}(\mathbf{P}) = n$ so it is invertible. We can compute

$$\mathbf{AP} = \mathbf{A}\left[\mathbf{x}_1|\mathbf{x}_2|\ldots|\mathbf{x}_n\right] = \left[\mathbf{Ax}_1|\mathbf{Ax}_2|\ldots|\mathbf{Ax}_n\right] = \left[\lambda_1\mathbf{x}_1|\lambda_2\mathbf{x}_2|\ldots|\lambda_n\mathbf{x}_n\right] = \mathbf{P}\begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}. \quad (2)$$

Multiplying by $\mathbf{P}^{-1}$ on the left yields

$$\mathbf{P}^{-1}\mathbf{AP} = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}, \quad (3)$$

so $\mathbf{A}$ is similar to a diagonal matrix. That proves the "if" direction of the theorem.

To show the "only if" direction, start with (3), where λ_i are some numbers, not necessarily eigenvalues, and all we know about $\mathbf{P}$ is that it is invertible. Multiplying by $\mathbf{P}$ on the left yields (2), which means the columns of $\mathbf{P}$ had to be eigenvectors of $\mathbf{A}$. Since $\mathbf{P}$ is invertible, these eigenvectors are linearly independent. □

Since the characteristic polynomial is degree n, the sum of the algebraic multiplicities of the eigenvalues is n. Section 6.5 Theorem 2 says that the geometric multiplicity of an eigenvalue is less than or equal to its algebraic multiplicity. Thus if any eigenvalue has geometric multiplicity less than its algebraic multiplicity, then the total of the geometric multiplicities will be less than n, so by Theorem 11 $\mathbf{A}$ is not diagonalizable.

Section 6.5 Theorem 3 says that eigenvectors from different eigenvalues are linearly independent, so combining linearly independent eigenvectors from different eigenvalues results in a linearly independent set. Thus, if every eigenvalue has geometric multiplicity equal to its algebraic multiplicity, then we can the basis for each and obtain n linearly independent eigenvectors.

Consequently, we have an algorithm for trying to diagonalize a matrix:

1. Find the eigenvalues.
2. For each eigenvalue:
   (a) Find a basis for its eigenspace.
   (b) If the geometric multiplicity (dimension of the eigenspace) is less that the algebraic multiplicity, then stop since the matrix is not diagonalizable.
3. Collect the bases for each eigenspace together. They will be linearly independent.
4. Form a matrix $\mathbf{P}$ whose columns are the vectors you collected.
5. Form the diagonal matrix $\mathbf{D}$ that has the eigenvalues on the diagonal, in the same order as the vectors in $\mathbf{P}$.

Notes:

- To form $\mathbf{D}$ we did not need the eigenvectors; we just had to check that the geometric and algebraic multiplicities of each eigenvalue are equal.

- To check that it worked, we should check $\mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$. Slightly easier is to check $\mathbf{P}\mathbf{D} = \mathbf{A}\mathbf{P}$; this is equivalent as long as $\mathbf{P}$ really is invertible.
- For an example see `https://en.wikipedia.org/wiki/Diagonalizable_matrix#How_to_diagonalize_a_matrix`

2.1 Example Uses

Suppose we found that $\mathbf{A}$ is similar to a diagonal matrix $\mathbf{D}$ and we know $\mathbf{P}$. Then $\mathbf{A}^k = \left(\mathbf{P}\mathbf{D}\mathbf{P}^{-1}\right)^k = \mathbf{P}\mathbf{D}^k\mathbf{P}^{-1}$. It is really easy to compute $\mathbf{D}^k$ since you just take each diagonal entry to the power k, so the diagonalization of $\mathbf{A}$ gives us a good way to compute $\mathbf{A}^k$. You can use this idea to evaluate polynomials of matrices as well, since $a\mathbf{A}^k + b\mathbf{A}^m = \mathbf{P}\left(a\mathbf{D}^k + b\mathbf{D}^m\right)\mathbf{P}^{-1}$.

Suppose you are given a set of eigenvalues $\{\lambda_i\}$ with corresponding linearly independent eigenvectors $\{\mathbf{x}_i\}$ and need to find $\mathbf{A}$. Make a diagonal matrix $\mathbf{D}$ with the eigenvalues on the diagonal. Make a matrix $\mathbf{P}$ with columns made up of the eigenvectors. Then $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ has those eigenvalues and eigenvectors.

2.2 Good Problem 6

For the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & -1 \\ 2 & 5 \end{bmatrix},$$

find a diagonal matrix $\mathbf{D}$ and an invertible matrix $\mathbf{P}$ such that $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$. Use this diagonalization to get formulas for each of the entries of $\mathbf{A}^k$.

3 Introduction to the Jordan Normal Form

If a matrix is not diagonalizable, we want to bring it as close to diagonal as we can and show how the diagonization failed. To do so we bring it to the Jordan normal form (also called Jordan canonical form).

Definition 12. *A* **Jordan block** *is a matrix that has an eigenvalue on the diagonal, 1 in all the entries just above the diagonal, and is zero otherwise, as in*

$$\begin{bmatrix} \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 \\ 0 & 0 & \lambda & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{bmatrix}. \tag{4}$$

Definition 13. *A* **Jordan normal form** *is a matrix that has Jordan blocks on the diagonal and is zero otherwise, as in*

$$\begin{bmatrix} \mathbf{J}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{J}_m \end{bmatrix}. \tag{5}$$

Theorem 14. *Every matrix is similar to a matrix in Jordan normal form.*

We will not try to prove this theorem in general, since you have to keep track of many possible cases. Notice that if the Jordan blocks are all 1×1 then the Jordan normal form is diagonal and the matrix is diagonalizable. When an eigenvalue has geometric multiplicity equal to its algebraic multiplicity k, then it gives rise to k Jordan blocks, each 1×1, and that part of the Jordan normal form is diagonal. For this class we will only consider the case when exactly one eigenvalue has geometric multiplicity less than its algebraic multiplicity.

The examples in Section 6.5 illustrate what can happen when an eigenvalue has algebraic multiplicity $k = 3 > 1$. In Example 3, the geometric multiplicity is 3, since there are 3 linearly independent eigenvectors. The matrix is already in its Jordan normal form, which has three 1×1 Jordan blocks and so is diagonal. In Example 1, the geometric multiplicity is 1, since there is only 1 linearly independent eigenvector. The matrix is already in its Jordan normal form, which has one 3×3 Jordan block. In Example 2, the geometric multiplicity is 2, since there are 2 linearly independent eigenvectors. The matrix is already in its Jordan normal form, which has one 2×2 Jordan block and one 1×1 Jordan block.

If the algebraic multiplicity is greater than 3, then more things could happen. For example, if $k = 4$ then we could get

- four 1×1 Jordan blocks,
- two 1×1 Jordan blocks and one 2×2 Jordan block,
- one 1×1 Jordan block and one 3×3 Jordan block, or
- two 2×2 Jordan blocks.

For this class we will only consider algebraic multiplicity at most 3.

We thus only need to figure out what to do when an eigenvalue does not have a full set of eigenvectors. We will go through the case when the algebraic multiplicity is 3 and the geometric multiplicity is 1. Let $\mathbf{x}_i$ be column i of $\mathbf{P}$. We want $\mathbf{AP} = \mathbf{PJ}$ and can compute

$$\mathbf{AP} = \mathbf{A}\left[\mathbf{x}_1|\mathbf{x}_2|\mathbf{x}_3\right] = \left[\mathbf{A}\mathbf{x}_1|\mathbf{A}\mathbf{x}_2|\mathbf{A}\mathbf{x}_3\right] \quad \text{and} \tag{6}$$

$$\mathbf{PJ} = \left[\mathbf{x}_1|\mathbf{x}_2|\mathbf{x}_3\right] \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix} = \left[\lambda\mathbf{x}_1|\mathbf{x}_1 + \lambda\mathbf{x}_2|\mathbf{x}_2 + \lambda\mathbf{x}_3\right]. \tag{7}$$

Equating these tells us

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x}_1 = \mathbf{0}, \quad (\mathbf{A} - \lambda\mathbf{I})\mathbf{x}_2 = \mathbf{x}_1 \quad \text{and} \quad (\mathbf{A} - \lambda\mathbf{I})\mathbf{x}_3 = \mathbf{x}_2. \tag{8}$$

Thus $\mathbf{x}_1$ is the ordinary eigenvector, $\mathbf{x}_2$ can be found from $\mathbf{x}_1$ by solving a linear system, and $\mathbf{x}_3$ can be found from $\mathbf{x}_2$ by solving a linear system.

For an example see `https://en.wikipedia.org/wiki/Jordan_normal_form#Example`

3.1 Example Uses

The example uses for the Jordan normal form are the same as those for diagonalization, only messier.

Letting $\mathbf{J}$ be the Jordan normal form of $\mathbf{A}$, we still have $\mathbf{A}^k = \mathbf{P}\mathbf{J}^k\mathbf{P}^{-1}$. The matrix $\mathbf{J}^k$ has powers of Jordan blocks on its diagonal. Since Jordan blocks bigger than 1×1 are not diagonal, it is harder to compute their powers. There is a general formula for their powers, but we will only state the 2×2 and 3×3 cases. (One of the WeBWork problems uses (9).)

Lemma 15.

$$\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}^k = \begin{bmatrix} \lambda^k & k\lambda^{k-1} \\ 0 & \lambda^k \end{bmatrix} \quad and \tag{9}$$

$$\begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}^k = \begin{bmatrix} \lambda^k & k\lambda^{k-1} & \frac{k(k-1)}{2}\lambda^{k-2} \\ 0 & \lambda^k & k\lambda^{k-1} \\ 0 & 0 & \lambda^k \end{bmatrix}. \tag{10}$$

Similar to the diagonal case, given the eigenvalues, eigenvectors, and extra vectors to make up for the lack of enough eigenvectors, we could construct the corresponding matrix $\mathbf{A}$. (However, we will not do this.)

3.2 Exercises

Find a matrix $\mathbf{P}$ that is invertible and a matrix $\mathbf{J}$ that is in Jordan normal form such that $\mathbf{A} = \mathbf{PJP}^{-1}$.

1.
$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 2 & 2 \end{bmatrix}$$

2.
$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 27 & -27 & 9 \end{bmatrix}$$

4 Norms

A norm is a way of measuring the size of an object such as a vector.

Definition 16. *A norm is a function yielding a real number such that*

1. $\|\mathbf{x}\| \geq 0$ *and* $\|\mathbf{x}\| = 0$ *implies* $\mathbf{x} = \mathbf{0}$,
2. $\|a\mathbf{x}\| = |a|\|\mathbf{x}\|$ *for any scalar* a, *and*
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

The third property is called the triangle inequality and means that the length of one side of a triangle cannot be more than the sum of the lengths of the other two sides.

For a vector with n real entries, we have already seen the norm

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}\,, \tag{11}$$

but there are many others. When we need to distinguish different norms, (11) is called the Euclidean norm or L^2 norm and is denoted $\|\mathbf{x}\|_2$.

We will consider two other norms.

Definition 17. *For a vector* $\mathbf{x}$ *with* n *real entries, the* L^∞ *norm is*

$$\|\mathbf{x}\|_\infty = \max\{|x_1|, |x_2|, \ldots, |x_n|\}\,. \tag{12}$$

Definition 18. *For a vector* $\mathbf{x}$ *with* n *real entries, the* L^1 *norm is*

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i| \,. \tag{13}$$

The choice of norm to use is often determined by the application. For example, if something must always be small or the plane will crash, then the L^∞ norm should be used; if you are measuring the distance between two addresses in Manhattan by taxi, then the L^1 norm should be used.

There are also norms on matrices. One way to get a matrix norm is to pretend it is a vector and use one of the vector norms. For example, if we use the vector L^2 norm, we get the Frobenius norm:

Definition 19. *For a* $n \times m$ *matrix* $\mathbf{A}$ *with real entries, the Frobenius norm is*

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2} \,. \tag{14}$$

A second way is to *induce* a matrix norm from a vector norm, by defining

$$\|\mathbf{A}\| = \max_{\mathbf{x}\neq 0} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\| \,. \tag{15}$$

We could choose any vector norm to use for $\|\mathbf{A}\mathbf{x}\|$ and any vector norm to use for $\|\mathbf{x}\|$ and get a matrix norm. Induced norms tell us the most $\mathbf{A}$ can stretch vectors. (If $\|\mathbf{A}\| < 1$ then $\mathbf{A}$ shrinks all vectors.) Induced norms can be easy or hard to compute. One can show

$$\|\mathbf{A}\|_1 = \max_{\mathbf{x}\neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_1}{\|\mathbf{x}\|_1} = \max_{\|\mathbf{x}\|_1=1} \|\mathbf{A}\mathbf{x}\|_1 = \max_{j=1,\ldots,m} \sum_{i=1}^{n} |a_{ij}| \,, \tag{16}$$

$$\|\mathbf{A}\|_\infty = \max_{\mathbf{x}\neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} = \max_{\|\mathbf{x}\|_\infty=1} \|\mathbf{A}\mathbf{x}\|_\infty = \max_{i=1,\ldots,n} \sum_{j=1}^{m} |a_{ij}| \,, \quad \text{and} \tag{17}$$

$$\|\mathbf{A}\|_2 = \max_{\mathbf{x}\neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2 = \max\left\{\sqrt{|\lambda|} \;:\; \lambda \text{ an eigenvalue of } \mathbf{A}^T\mathbf{A}\right\} . \tag{18}$$

4.1 Exercises

1. Show that the L^1 norm in (13) satisfies the conditions in Definition 16 to be a norm.
2. For $\mathbf{x} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$, compute
   (a) $\|\mathbf{x}\|_1$,
   (b) $\|\mathbf{x}\|_2$, and
   (c) $\|\mathbf{x}\|_\infty$.
3. For $\mathbf{A} = \begin{bmatrix} 1 & -2 \\ 3 & 4 \end{bmatrix}$, compute
   (a) $\|\mathbf{A}\|_F$,
   (b) $\|\mathbf{A}\|_1$,
   (c) $\|\mathbf{A}\|_2$, and
   (d) $\|\mathbf{A}\|_\infty$.