COORDINATES, EIGENVALUES, AND EIGENVECTORS

1. Review of coordinates

1.1. **The standard basis.** The *standard basis* for $\mathbb{R}^n$ is the basis $\mathcal{E} = \{e_1, e_2, \dots e_n\}$, where e_j is the vector with 1 in the j-th position and zeros elsewhere.

1.2. **Coordinates with respect to a basis.** Let $\mathcal{B} = \{v_1, v_2, \dots v_n\}$ be a basis for $\mathbb{R}^n$. We know that for each vector $x \in \mathbb{R}^n$ there are unique scalar "coordinates" $c_1, c_2, \dots c_n$ such that

$$x = c_1 v_1 + c_2 v_2 + \dots + c_n v_n \tag{1.1}$$

We call the vector $c \in \mathbb{R}^n$ having coordinates $c_1, c_2 \dots c_n$ the "coordinate vector" of x with respect to the basis $\mathcal{B}$, and denote it by $[x]_\mathcal{B}$.

1.3. **First examples.**

(a) $[x]_\mathcal{E} = x$ for each $x \in \mathbb{R}^n$.

(b) For any basis $\mathcal{B} = \{v_1, v_2, \dots v_n\}$ of $\mathbb{R}^n$:

$$[v_j]_\mathcal{B} = e_j (1 \le j \le n).$$

(c) $[0]_\mathcal{B} = 0$ for any basis $\mathcal{B}$ (the zero vector has the same representation with respect to any basis.

1.4. **Matrix calculation of coordinates.** Letting $\mathcal{B} = \{v_1, v_2, \dots v_n\}$ be a basis for $\mathbb{R}^n$ we can rewrite equation (1.1) in matrix form as

$$x = V[x]_\mathcal{B} \tag{1.2}$$

where V is the matrix having the vector v_j for its j-th column, i.e.

$$V = [v_1, v_2, \dots v_n].$$

Date: March 6, 2005.

V is invertible because its columns are linearly independent (see, for example, Summary 3.2.9(v) of our textbook, pp. 119–120), so we can compute the $\mathcal{B}$ coordinate representation of any vector $x \in \mathbb{R}^n$ by solving (1.2) for $[x]_{\mathcal{B}}$:

$$[x]_{\mathcal{B}} = V^{-1}x \tag{1.3}$$

For example, if in $\mathbb{R}^2$ we take $\mathcal{B} = \{v_1, v_2\}$ to be our basis, where $v_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $v_2 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$, then

$$V = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix} \quad \text{and} \quad V^{-1} = \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix}.$$

Suppose, for instance, that $x = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Then

$$[x]_{\mathcal{B}} = V^{-1}x = \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -7 \\ 4 \end{bmatrix}$$

More generally, if $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is any vector in $\mathbb{R}^n$ then the same calculation shows that

$$[x]_{\mathcal{B}} = \begin{bmatrix} -5x_1 + 2x_2 \\ 3x_1 - x_2 \end{bmatrix}.$$

1.5. **The matrix of a linear transformation with respect to a basis.** Recall that each $n \times n$ matrix induces a *linear transformation* $T : \mathbb{R}^n \to \mathbb{R}^n$ by means of matrix multiplication:

$$T(x) = Ax \qquad (x \in \mathbb{R}^n). \tag{1.4}$$

We refer to A as the "standard matrix" for T. The lesson of what's to follow is that it's not always the best matrix for T!

For a basis $\mathcal{B} = \{v_1, v_2, \dots v_n\}$ for $\mathbb{R}^n$ we *define*

$$[T]_{\mathcal{B}} = [[T(v_1)]_{\mathcal{B}}, [T(v_2)]_{\mathcal{B}} \dots [T(v_n)]_{\mathcal{B}}], \tag{1.5}$$

i.e. $[T]_{\mathcal{B}}$ is the matrix that has the $\mathcal{B}$-coordinate representation of $T(v_j)$ as its j-th column $(1 \le j \le n)$. We call $[T]_{\mathcal{B}}$ the *matrix of T with respect to the basis $\mathcal{B}$.*

1.6. **Example: $\mathcal{B} = \mathcal{E}$, the standard basis.** In this case $[T]_{\mathcal{B}} = [Ae_1, Ae_2, \dots Ae_n]$, the "standard matrix" for T.

1.7. **Geometric examples in $\mathbb{R}^2$.** In this next group of examples, L is a line through the origin in the direction of a non-zero vector $v \in \mathbb{R}^2$, and w is a non-zero vector perpendicular to v. We take as our basis for $\mathbb{R}^2$, not the standard one, but instead $\mathcal{B} = \{v, w\}$.

(a) Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is the transformation of *orthogonal projection onto L.* Then $T(v) = v$ and $T(w) = 0$, so

$$[T]_{\mathcal{B}} = [[T(v)]_{\mathcal{B}}, [T(w)]_{\mathcal{B}}] = [[v]_{\mathcal{B}}, [0]_{\mathcal{B}}] = [e_1, 0] = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

where we use the fact that $[v]_{\mathcal{B}} = e_1$ and $[0]_{\mathcal{B}} = 0$ (see §1.3 (b) and (c)).

Compare this simple and informative matrix for T with the one we got when we first introduced orthogonal projection. Then we took v to be a unit vector with coordinates u_1 and u_2, and did this computation:

$$\begin{aligned} T(x) &= (x \cdot v)v = (x_1u_2 + x_2u_2)\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \\ &= \begin{bmatrix} u_1^2x_1 + u_1u_2x_2 \\ u_1u_2x_1 + u_2^2x_2 \end{bmatrix} = \begin{bmatrix} u_1^2 & u_1u_2 \\ u_1u_2 & u_2^2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \end{aligned}$$

so the "standard matrix" for the linear transformation of orthogonal projection onto L is the completely uninformative one

$$A = \begin{bmatrix} u_1^2 & u_1u_2 \\ u_1u_2 & u_2^2 \end{bmatrix}.$$

(b) Suppose T is *reflection in L.* Then $T(v) = v$ as before, but now $T(w) = -w$. Thus

$$[T]_{\mathcal{B}} = [[T(v)]_{\mathcal{B}}, [T(w)]_{\mathcal{B}}] = [[v]_{\mathcal{B}}, [-w]_{\mathcal{B}}] = [e_1, -e_2] = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Note that in this computation we used the fact that $[-w]_{\mathcal{B}} = -[w]_{\mathcal{B}}$ ("linearity of coordinates"—Fact 3.4.2, page 139 of our textbook) and that $[v]_{\mathcal{B}} = e_1$ and $[w]_{\mathcal{B}} = e_2$ (§1.3 (b)).

(c) Suppose T is a *shear* in the direction of L, i.e. that $T(v) = v$ and $T(w) = kv + w$ for some fixed real number k. Then using linearity of coordinates and §1.3(b) we see that

$$[T]_{\mathcal{B}} = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$$

which is the same as the *standard matrix* for a *horizontal* shear.

1.8. **Using $[T]_{\mathcal{B}}$ to compute $T(x)$.** The utility of the matrix of a linear transformation T with respect to a basis $\mathcal{B}$ lies in the following formula:

$$[T]_{\mathcal{B}}[x]_{\mathcal{B}} = [T(x)]_{\mathcal{B}} \qquad (x \in \mathbb{R}^n), \tag{1.6}$$

which says that, if you're willing to give up the standard basis fro $\mathbb{R}^n$, and do all your calculations relative to the basis $\mathcal{B}$, then you can use the matrix of T with respect to $\mathcal{B}$ in the same way you used to use the standard matrix.

Proof of formula (1.6): For $x \in \mathbb{R}^n$ we have the unique representation

$$x = c_1v_1 + c_2v_2 + \ldots + c_nv_n$$

where $c_1, c_2, \ldots, c_n$ are the (standard) coordinates of $[x]_{\mathcal{B}}$. By linearity of T:

$$T(x) = c_1T(v_1) + c_2T(v_2) + \ldots + c_nT(v_n),$$

so by linearity of coordinates:

$$[T(x)]_{\mathcal{B}} = c_1[T(v_1)]_{\mathcal{B}} + c_2[T(v_2)]_{\mathcal{B}} + \ldots + c_n[T(v_n)]_{\mathcal{B}},$$

But the expression on the right is just the matrix product $[T]_{\mathcal{B}}[x]_{\mathcal{B}}$, so we are done.

2. Eigenvalues and Eigenvectors

2.1. **Definition.** Suppose A is an $n \times n$ matrix and T the associated linear transformation of $\mathbb{R}^n$ (so $T(x) = Ax$ for every $x \in \mathbb{R}^n$). We call a real number λ an *eigenvalue* of T (or of A) whenever there is a non-zero vector $x \in \mathbb{R}^n$ such that $T(x) = \lambda x$. The vector x is called an *eigenvector* of T (or of A) corresponding to the eigenvalue λ.

2.2. **Examples.**

(a) Look back at the examples of §1.7, where L is a line through the origin in the direction of a non-zero vector v, and w is a non-zero vector perpendicular to v. If $T : \mathbb{R}^2 \to \mathbb{R}^2$ is orthogonal projection onto L, then v and w are eigenvectors of T corresponding respectively to eigenvalues 1 and 0. If T is reflection in L then v and w are still eigenvectors, but now the corresponding eigenvalues are 1 and -1. If T is a shear along L then v is an eigenvector with eigenvalue 1, but w is not an eigenvector.

(b) (see Problem 25, page 146). Suppose $A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$ and $v = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $w = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Then one computes that $Av = 5v$ and $Aw = -w$, so v is an eigenvector for A corresponding to the eigenvalue 5 and w is an eigenvector corresponding to the eigenvalue -1. Note that, with respect to the basis $\mathcal{B} = \{v, w\}$ of $\mathbb{R}^2$, the matrix for the linear transformation induced by A is just $\begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix}$.

(c) (see Problem 29 of §3.4). Here T is the linear transformation of $\mathbb{R}^3$ induced by the matrix

$$A = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -2 & 2 \\ 3 & -9 & 6 \end{bmatrix}$$

We are given a basis $\mathcal{B} = \{v_1, v_2, v_3\}$ for $\mathbb{R}^3$, where

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \text{and } v_3 = \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix}.$$

You can check easily that each of these basis vectors is an eigenvector: $Av_1 = 0$ (eigenvalue zero), $Av_2 = v_2$ (eigenvalue 1) and $Av_3 = 2v_3$ (eigenvalue 2), and that therefore the matrix of the associated linear transformation T with respect to the basis $\mathcal{B}$ is

$$[T]_{\mathcal{B}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

In each of the examples of the previous two sections, whenever we had a linear transformation T of $\mathbb{R}^n$ that bequeathed to $\mathbb{R}^n$ a basis $\mathcal{B}$ of eigenvectors, the matrix of T with respect to B turned out to be *diagonal*, i.e. all entries zero off the main diagonal.

2.3. **Notation.** The $n \times n$ diagonal matrix with entry λ_j in the j-th diagonal position will be denoted diag $\{\lambda_1, \lambda_2, \dots \lambda_n\}$.

Thus, for example, the 3×3 identity matrix is diag $\{1, 1, 1\}$, while in §2.2(c), $[T]_{\mathcal{B}} =$ diag $\{0, 1, 2\}$.

Diagonal matrices are the easiest ones for which to find eigenvalues and eigenvectors:

2.4. **Proposition.** *If $A =$ diag $\{\lambda_1, \lambda_2, \dots \lambda_n\}$ then the standard unit vector e_j is an eigenvector for the eigenvalue λ_j $\quad (1 \le j \le n)$.*

Proof. Our hypothesis is that $A = [\lambda_1 e_1, \lambda_2 e_2, \ldots, \lambda_n e_n]$, i.e. that A is the matrix with $\lambda_j e_j$ as its j-th column ($1 \le j \le n$). But Ae_j is the j-th column of A, so $Ae_j = \lambda_j e_j$ for each j between 1 and n. $\square$

2.5. **Exercise.** Prove the converse: *If an $n \times n$ matrix A has the standard basis for $\mathbb{R}^n$ as eigenvectors, then A is diagonal.*

2.6. **The Diagonalization Theorem.** *For a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$, and a basis $\mathcal{B}$ for $\mathbb{R}^n$, the following two statements are equivalent (each implies the other):*

(a) *$\mathcal{B}$ consists of eigenvectors of T.*

(b) *$[T]_{\mathcal{B}}$ is a diagonal matrix.*

Proof. (a) $\Rightarrow$ (b): Suppose $\mathcal{B} = \{v_1, v_2, \ldots v_n\}$ consists of eigenvectors of T, so that there are real numbers λ_j ($1 \le j \le n$) such that

$$T(v_j) = \lambda_j v_j \qquad (j = 1, 2, \ldots, n).$$

We must show that $[T]_{\mathcal{B}}$ is diagonal. Starting with the definition of $[T]_{\mathcal{B}}$, we compute:

$$\begin{aligned} [T]_{\mathcal{B}} &= [[Tv_1]_{\mathcal{B}}, [Tv_2]_{\mathcal{B}}, \ldots, [Tv_n]_{\mathcal{B}}] \\ &= [[\lambda_1 v_1]_{\mathcal{B}}, [\lambda_2 v_2]_{\mathcal{B}}, \ldots, [\lambda_n v_n]_{\mathcal{B}}] \\ &= [\lambda_1 [v_1]_{\mathcal{B}}, \lambda_2 [v_2]_{\mathcal{B}}, \ldots, \lambda_n [v_n]_{\mathcal{B}}] \\ &= [\lambda_1 e_1, \lambda_2 e_2, \ldots, \lambda_n e_n] \\ &= \operatorname{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\} \end{aligned}$$

where, in this computation we used successively: the definition of $[T]_{\mathcal{B}}$ (line 1), the fact that $T(v_j) = \lambda_j v_j$ (line 2), linearity of coordinates (line 3), and finally the fact that $[v_j]_{\mathcal{B}} = e_j$ (line 4).

(b) $\Rightarrow$ (a): Now suppose $\mathcal{B} = \{v_1, v_2, \ldots v_n\}$ is a basis for $\mathbb{R}^n$, and that $[T]_{\mathcal{B}} = \operatorname{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$. We must show that each v_j is an eigenvector of T. In fact, we'll see that $Tv_j = \lambda_j v_j$ for $1 \le j \le n$. Fix such a j, and note that the j-th column of $[T]_{\mathcal{B}}$ is just $\lambda_j e_j$, i.e.,

$$[T]_{\mathcal{B}} e_j = \lambda_j e_j \tag{2.1}$$

Now the right-hand side of (2.1) is $\lambda_j[v_j]_{\mathcal{B}}$ which, by linearity of coordinates, is $[\lambda_j v_j]_{\mathcal{B}}$. Similarly, the left-hand side of (2.1) is $[T]_{\mathcal{B}}[v_j]_{\mathcal{B}}$ which, by (1.6), is $[Tv_j]_{\mathcal{B}}$. Thus

$$[Tv_j]_{\mathcal{B}} = [\lambda_j v_j]_{\mathcal{B}},$$

i.e., the both Tv_j and $\lambda_j v_j$ have the same $\mathcal{B}$-coordinate vectors. Thus $Tv_j = \lambda_j v_j$, as desired. $\square$

2.7. **Definition.** Call an $n \times n$ matrix *diagonalizable* if $\mathbb{R}^n$ has a basis consisting of eigenvectors of this matrix.

Warning: Not all square matrices are diagonalizable! For example it's geometrically clear that shear transfomations of $\mathbb{R}^2$ have (up to multiplication by a non-zero constant) just a single eigenvector, while rotations of $\mathbb{R}^2$ through any angle not an integer multiple of π have no eigenvectors at all! We'll see how to prove this kind of result"analytically" (i.e. without reference to pictures) in the next section.

3. Efficient eigenvalue computation for 2×2 matrices

We'll base our computations for 2×2 matrices on the notion of *determinant*.

3.1. **Definition.** The *determinant* of the matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is:

$$\det(A) = ad - bc$$

The key to all the work of this section is:

3.2. **Theorem.** *For A a 2×2 matrix, the system of linear equations $Ax = 0$ has non-trivial solution if and only if* $\det(A) = 0$.

Proof. In scalar form the system of equations is:

$$\begin{aligned} ax_1 + bx_2 &= 0 \\ cx_1 + dx_2 &= 0 \end{aligned}$$

Suppose first that neither a nor c is zero. Then we can replace the second equation by c times the first minus a times the second, yielding the equivalent system:

$$\begin{aligned} ax_1 + bx_2 &= 0 \\ (ad - bc)x_2 &= 0 \end{aligned}$$

which has a nontrivial solution if and only if $ad - bc = 0$.

Suppose $a = 0$. Then $\det(A) = -bc$, which is zero precisely when either b or c is zero. If $b = 0$ then we are back to one equation in two unknowns, so there are nontrivial (infinitely many) solutions. If $b \neq 0$ then any solution must have $x_2 = 0$, whereupon the second equation implies $cx_1 = 0$. If $c = 0$ (i.e. $\det(A) = 0$) then there are infinitely many solutions: $x_2 = 0$ and x_1 arbitrary. If, however $c \neq 0$ (i.e., $\det(A) \neq 0$) then x_1 is also required to be zero, hence the only solution is the trivial one. $\square$

3.3. **Corollary.** *The eigenvalues of a* 2×2 *matrix* A *are the real numbers* λ *that satisfy its "characteristic equation"*

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

Proof. This follows immediately from the above Theorem and the definition of eigenvalue. More precisely, to say a real number λ is an eigenvalue of A means that there is a non-zero vector $x \in \mathbb{R}^2$ such that:

$$Ax = \lambda x \quad \text{i.e.} \quad Ax - \lambda x = 0 \quad \text{i.e.} \quad (A - \lambda I)x = 0$$

Thus λ is an eigenvector of A if and only if the homogeneous system

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

has a nontrivial solution, which, by Theorem 3.2, happens if and only if $\det(A - \lambda I) = 0$. $\square$

3.4. **Example.** Show that the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$$

is diagonalizable.

Proof. Of course we know this already from §2.2(b) where the eigenvectors and eigenvalues of A were handed to us. But suppose they weren't. Then we write the characteristic equation:

$$\begin{aligned} 0 &= \det(A - \lambda I) = \det(\begin{bmatrix} 1-\lambda & 2 \\ 4 & 3-\lambda \end{bmatrix}) \\ &= (1-\lambda)(3-\lambda) - 8 = \lambda^2 - 4\lambda - 5 \\ &= (\lambda - 5)(\lambda + 1) \end{aligned}$$

so that $\lambda = 5$ and $\lambda = -1$ are the solutions, and so these are the eigenvalues of A.

Now for $\lambda = 5$,

$$A - \lambda I = A - 5I = \begin{bmatrix} -4 & 2 \\ 4 & -2 \end{bmatrix}$$

and the system $(A - 5I)x = 0$ is easily seen to have solutions $x =$ any scalar multiple of $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$. In particular, $v_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is the unique (up to multiplication by a non-zero scalar) eigenvector for the eigenvalue 5.

I leave it to you to show by the same method that $v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ is the "unique" eigenvector for the eigenvalue -1.

It's easy to check that $\mathcal{B} = \{v_1, v_2\}$ is a basis for $\mathbb{R}^2$, so A is diagonalizable, and the matrix of its linear transformation, with respect to the basis $\mathcal{B}$ is diag $\{5, -1\}$. $\square$

3.5. **Example: a non-diagonalizable matrix.** Let $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. The characteristic equation for this matrix is $\lambda^2 = 0$, which has only $\lambda = 0$ as its solution. The corresponding eigenvectors are the nontrivial solutions to the equation $Ax = 0$ (i.e. the non-zero vectors in the kernel of A), and these all the non-zero scalar multiples of the standard unit vector e_1. Thus there aren't enough eigenvectors here to make a basis of $\mathbb{R}^2$, so A is not diagonalizable.

4. Matrix of $[T]_{\mathcal{B}}$ vs. standard matrix of T

Suppose T is a linear transformation of $\mathbb{R}^n$ whows "standard matrix" is A, i.e.

$$T(x) = Ax \qquad (x \in \mathbb{R}^n).$$

Suppose $\mathcal{B} = \{v_1, v_2, \ldots v_n\}$ is a basis for $\mathbb{R}^n$. It's time to get serious about the relationship between A and $[T]_{\mathcal{B}}$.

Recall that in §1.8 we derived equation (1.3):

$$[x]_{\mathcal{B}} = V^{-1}x$$

where $V = [v_1, v_2, \ldots, v_n]$ is the matrix having the vector v_j as its j-th column. Recall also that the invertibility of V derives from the fact that its columns are linearly independent.

Thus we may compute:

$$
\begin{aligned}
[T]_{\mathcal{B}} &= [[T(v_1)]_{\mathcal{B}}, [T(v_2)]_{\mathcal{B}}, \ldots, [T(v_n)]_{\mathcal{B}}] \\
&= [[Av_1]_{\mathcal{B}}, [Av_2]_{\mathcal{B}}, \ldots, [Av_n]_{\mathcal{B}}] \\
&= \left[V^{-1}Av_1, V^{-1}Av_2, \ldots, V^{-1}Av_n\right] \\
&= V^{-1}A\,[v_1, v_2, \ldots, v_n] \\
&= V^{-1}AV
\end{aligned}
$$

where the first line is the definition of $[T]_{\mathcal{B}}$, the second uses the definition of T in terms of A, the third uses (1.3), and the fourth is the definition of matrix multiplication.

We have proved:

4.1. **Theorem** (Fact 3.4.4, Page 144 of our text). *Suppose $\mathcal{B} = \{v_1, v_2, \ldots v_n\}$ is a basis for $\mathbb{R}^n$ and T a linear transformation of $\mathbb{R}^n$ with standard matrix A. Then $[T]_{\mathcal{B}} = V^{-1}AV$, where V is the matrix whose j-th column is v_j $(1 \le j \le n)$.*

5. Application: Powers of matrices

Theorem 4.1 is of great utility in computing large powers of matrices—a calculation that would be difficult to do by hand in many circumstances. The method depends on two crucial facts.

5.1. **Theorem.** *If A, B and V are $n \times n$ matrices, and V is invertible, then for each positive integer p:*

$$(VBV^{-1})^p = VB^pV^{-1}.$$

Proof. We use the method of Mathematical Induction, which asserts that to prove the theorem we need only check that it's true for $p = 1$, and then show that:

(*) *Result true for an integer $p \ge 1$ $\Rightarrow$ Result true for $p + 1$.*

The point is that once you've established the case $p = 1$ and proved (*), then (*) gives you for free the case $p = 2$, and then the case $p = 3$, and so on forever.

For our result, the case $p = 1$ is obviously true, so suppose we have proved the theorem for some $p \geq 1$. We must show that it's true for $p+1$; to this end write

$$\begin{aligned}(VBV^{-1})^{p+1} &= (VBV^{-1})^p(VBV^{-1}) \\ &= (VB^pV^{-1})(VBV^{-1}) \\ &= (VB^p)(V^{-1}V)(BV^{-1}) \\ &= (VB^p)I(BV^{-1}) \\ &= VB^{p+1}V^{-1}\end{aligned}$$

where the second line follows from our "induction hypothesis" that the theorem is true for p, the third line follows from the associate law of matrix multiplication, and the others are clear. □

The next result asserts that for a *diagonal* matrix you can compute a power by taking the corresponding power of each element of the matrix.

5.2. **Theorem.** *For any positive integer p,*

$$\operatorname{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}^p = \operatorname{diag}\{\lambda_1^p, \lambda_2^p, \ldots, \lambda_n^p\}$$

Proof. I leave this one mostly as an exercise. First check that

$$\operatorname{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}\operatorname{diag}\{\mu_1, \mu_2, \ldots, \mu_n\} = \operatorname{diag}\{\lambda_1\mu_1, \lambda_2\mu_2, \ldots, \lambda_n\mu_n\}$$

The result about powers then follows quickly from this by induction. □

5.3. **Example.** Compute $\begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}^{15}$.

Solution. Recall that we've already found that the vectors $v_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ are eigenvectors for $A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$ (for eigenvalues 5 and -1 respectively) and that $\mathcal{B} = \{v_1, v_2\}$ is a basis for $\mathbb{R}^2$. Thus, according to Theorem 4.1 above,

$$\begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix} = [T]_{\mathcal{B}} = V^{-1}AV, \qquad \text{where} \qquad V = [v_1, v_2] = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}.$$

Thus

$$A = V \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix} V^{-1},$$

so successive application of Theorem 5.1 and Theorem 5.2 yield,

$$A^{15} = V \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix}^{15} V^{-1} = V \begin{bmatrix} 5^{15} & 0 \\ 0 & (-1)^{15} \end{bmatrix} V^{-1}. \tag{5.1}$$

A little calculation shows that $V^{-1} = \begin{bmatrix} 1/3 & 1/3 \\ 2/3 & -1/3 \end{bmatrix}$, after which (5.1) and a little more calculation (and the fact that $(-1)^{15} = -1$) yield

$$A^{15} = \frac{1}{3} \begin{bmatrix} 5^{15} - 1 & 2^{15} + 1 \\ 2^{15} + 1 & 4^{15} - 1 \end{bmatrix}.$$

5.4. **Example: A dynamical system.** Neighboring countries Elbonia and Pilikia are undergoing a relatively peaceful period in their histories. Each is reducing its annual defense budget, but because of nervousness about the others intentions, each adds back in to its budget a percentage of its neighbors expenditures.

More precisely, Pilikia reduces its budget by 25%, but adds back in 25% of Elbonia's budget, while Elbonia reduces its budget by a whopping 75%, but adds back in 75% of Pilikia's defense budget.

Let $P(t)$ denote Pilikia's defense budget for year t, and $E(t)$ Elbonia's (in 2005 Euros times one billion). Suppose initially $P(0) = 3$ and $E(0) = 2$. Determine the behavior of $P(t)$ and $E(t)$ as $t \to \infty$.

Solution The vector $x(t) = \begin{bmatrix} P(t) \\ E(t) \end{bmatrix}$ describes the state of the system at year t. We have (with amounts calculated in billions of Euros)

$$\begin{aligned} P(t+1) &= .75P(t) + .25E(t) \\ E(t+1) &= .25P(t) + .75E(t) \end{aligned}$$

i.e.,

$$x(t+1) = Ax(t) \qquad \text{where} \qquad A = \begin{bmatrix} .75 & .25 \\ .25 & .75 \end{bmatrix} \tag{5.2}$$

Thus defense expenditures for both countries will be: $x(1) = Ax(0)$ in year one, $x(2) = Ax(1) = A^2x(0)$ in year two, and more generally in year n they will be

$$x(n) = Ax(n-1) = A^2x(n-2) \cdots = A^n x(0).$$

So it all comes down to computing A^n.

For this I leave it to you to check that the eigenvalues of A are 1 and 1/2, and the corresponding eigenvectors are

$$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ for } \lambda = 1 \quad \text{and} \quad v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \text{ for } \lambda = 1/2.$$

Now $\mathcal{B} = \{v_1, v_2\}$ is a basis for $\mathbb{R}^2$ relative to which the linear transformation T whose standard matrix is A has a diagonal matrix:

$$[T]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix} = V^{-1}AV, \qquad \text{where} \qquad V = [v_1, v_2] = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Thus for each positive integer n,

$$\begin{bmatrix} 1^n & 0 \\ 0 & (1/2)^n \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix}^n = V^{-1}A^nV,$$

hence solving for A^n:

$$A^n = V \begin{bmatrix} 1^n & 0 \\ 0 & (1/2)^n \end{bmatrix} V^{-1}.$$

Now one calculates easily that

$$V^{-1} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}$$

from which follows

$$A^n = \frac{1}{2} \begin{bmatrix} 1 + 2^{-n} & 1 - 2^{-n} \\ 1 - 2^{-n} & 1 + 2^{-n} \end{bmatrix}.$$

Thus

$$x(n) = A^n[x(0)] = \frac{1}{2} \begin{bmatrix} (1 + 2^{-n}) & 1 - 2^{-n} \\ 1 - 2^{-n} & 1 + 2^{-n} \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 5 + 2^{-n} \\ 5 - 2^{-n} \end{bmatrix}.$$

Thus $x(n) \to \begin{bmatrix} 5/2 \\ 5/2 \end{bmatrix}$ as $n \to \infty$, i.e. both nations tend toward a yearly defense budget of 2.5 billion Euros.

6. Exercises

I. §7.1 #1–4, 8, 9, 12, 36, 38

II. §7.2 #28; this will be part of a future writing assignment.

III. §7.3 #40, 42 These will be part of a future writing assignment. §5.4 of these notes works out a special case of #40.