CISC 271 Class 1

Introduction, With a Review of Eigenvalues and Eigenvectors

Text Correspondence: §1.1–2.5, §6.1

Main Concepts:

- *Matrix: an ordered set of vectors*
- *Linear transformation: summation of vectors*
- *Characteristic polynomial and characteristic equation*
- *Eigenvalue: a root of $F(\lambda) = 0$*
- *Eigenvector: mapped, by a matrix, to a multiple of itself*

Sample Problem, Machine Inference: From a data matrix A and a weight vector $\vec{x}$, infer a data vector as $\vec{y} = A\vec{x}$

In this course, we will use linear algebra as a language for describing how to perform basic analysis of structured data. Fundamental concepts that we will assume as known include:

- Vectors as members of vector spaces
- Spanning sets and basis sets of a vector space
- Vector norms, especially the Euclidean norm $\|\cdot\|_2$
- Transpose of a vector
- Orthogonal vectors, relating norms to transposes
- Partitioning a matrix into blocks
- A matrix as an ordered list of vectors
- A linear equation as a matrix-vector equation
- A system of linear equations as a matrix-vector equation
- Representing a linear equation as an augmented matrix
- The reduced row echelon form (RREF) of a matrix
- The null space, or kernel, of a matrix
- The column span of a matrix
- Using the RREF to find a basis set from a matrix
- The determinant of a matrix
- For a nonsingular matrix, the inverse of a matrix

- Elementary properties of eigenvalues and eigenvectors

One ordinary understanding of the equation $\vec{y} = A\vec{x}$ is that the matrix A is a linear operator, with the weight vector $\vec{x}$ being an operand that is linearly mapped to the data vector $\vec{y}$. Another point of view – which was prominently formulated by Albert Einstein – is that the weight vector $\vec{x}$ is a linear operator that maps the data matrix A to the data vector $\vec{y}$. This alternate point of view will be very useful when we examine eigenvectors and principal components so let us now see why it is so.

Consider the 2×2 case, which is small enough to be quickly worked by hand and nontrivial. The linear equation is simply

$$\vec{c} = A\vec{w} \tag{1.1}$$

Working out the individual entries, Equation 1.1 is

$$\begin{aligned}\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\ &= \begin{bmatrix} a_{11}w_1 + a_{12}w_2 \\ a_{21}w_1 + a_{22}w_2 \end{bmatrix} \end{aligned} \tag{1.2}$$

We can arrive at the alternative formulation in two ways. One is to *partition* the data matrix in Equation 1.1 into *blocks* of column vectors, and write the weight vector in components. The data matrix can be partitioned into two column vectors, $\vec{a}_1$ and $\vec{a}_2$, so that

$$\vec{y} = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \tag{1.3}$$

One of the strengths of partitioning is that we can deal with the blocks in A and $\vec{w}$ as distinct mathematical entities. The first block of A, which is $\vec{a}_1$, is multiplied by the first block of $\vec{w}$, which is the weight w_1, to give $w_1\vec{a}_1$; likewise the second block of A, which is $\vec{a}_2$, is multiplied by the second block of $\vec{w}$, which is the weight w_2, to give $w_2\vec{a}_2$. The vector $\vec{y}$ is the sum of these block multiplications so

$$\vec{c} = w_1\vec{a}_1 + w_2\vec{a}_2$$

Another way to arrive at the same result is to carefully observe the component-by-component arrangement in Equation 1.2. This can be successively re-written as

$$\begin{aligned}
\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} &= \begin{bmatrix} a_{11}w_1 + a_{12}w_2 \\ a_{21}w_1 + a_{22}w_2 \end{bmatrix} \\
&= \begin{bmatrix} w_1 a_{11} + w_2 a_{12} \\ w_1 a_{21} + w_2 a_{22} \end{bmatrix} \\
&= w_1 \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + w_2 \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} \\
&= w_1 \vec{a}_1 + w_2 \vec{a}_2 \qquad (1.4)
\end{aligned}$$

A third way more closely follows Einstein's method (he did not have the matrix-vector notation that we now take for granted). The first entry of $\vec{c}$, which is w_1, can be written as a sum so that

$$c_1 = \sum_{j=1}^{2} w_j a_{1j}$$

A less than rigorous way of using the Einstein summation convention is to (a) omit the formal summation when the number of terms is well understood from context, and (2) to have the summation *implied* by the repeated use of an index. This would represent each resultant entry of our example matrix-vector product as

$$c_i = w_j a_{ij} \qquad (1.5)$$

In Equation 1.5, it is both an advantage and a disadvantage to not know ahead of time which is the operator and which is the operand. We will not use this notation further in this course but later courses may rely heavily on the notation.

These ideas can readily be generalized to a data vector $\vec{c} \in \mathbb{R}^m$ with m entries, a weight vector $\vec{x} \in \mathbb{R}^n$ with n entries, and a data matrix $A \in \mathbb{R}^{m \times n}$ with $m \times n$ entries. The matrix-vector product can be written as

$$\begin{aligned}
\vec{c} &= A\vec{w} \\
&= w_1 \vec{a}_1 + w_2 \vec{a}_2 + \cdots + w_n \vec{a}_n \\
&= \sum_{j=1}^{n} w_j \vec{a}_j \qquad (1.6)
\end{aligned}$$

As preparation for the rest of this course, a student should be thoroughly familiar with basic properties and conventions in linear algebra. The textbook is the best source that the instructor has found for learning linear algebra and as a reference text.

1.1 Eigenvalues and Eigenvectors

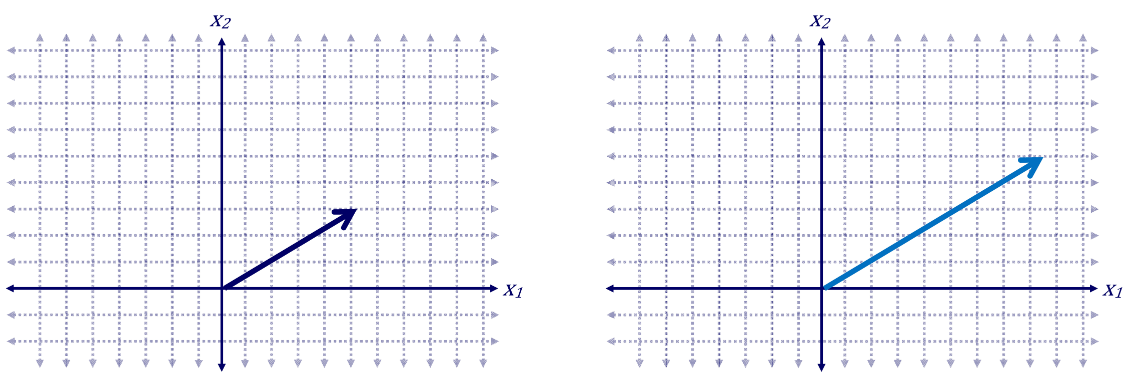
Eigenvalues and eigenvectors are useful in describing and analyzing many systems, both physical and mathematical. The definition of an eigenvalue λ and eigenvector $\vec{v}$, of a given square matrix $A \in \mathbb{R}^{n \times n}$, is that they satisfy both the equation and the restriction

$$A\vec{v} = \lambda\vec{v} \quad \text{with} \quad \vec{v} \neq \vec{0} \tag{1.7}$$

An eigenvector is special because its direction is not changed when it is transformed by the matrix A; its magnitude is scaled by its eigenvalue. The condition attached to Equation 1.7 is because the zero vector is always transformed linearly to the zero vector; we are only interested in the *nontrivial* solutions.

There are abundant examples and applications of eigenvalues available from other sources. For example, the extra notes for this class describe how one physical system can be described using eigenvalues. The example uses two masses connected with springs, forced to move sinusoidally. The eigenvalues correspond to stable frequencies that cause the masses to oscillate in a steady manner.

We will examine the basic properties of a corresponding eigenvalue and eigenvector, both algebraically and graphically. The first, elementary, observation from Equation 1.7 is that an eigenvector $\vec{v}$ of a matrix A is transformed to a multiple of itself: the associated eigenvalue, λ, acts as a scale factor. A 2D example of this is depicted in Figure 1.1.



(A) **(B)**

Figure 1.1: An eigenvector is transformed by a matrix as a change of its magnitude. (A) The original eigenvector $\vec{v}$. (B) The vector, scaled by the corresponding eigenvalue, becomes $\lambda\vec{v}$. The direction of the vector is unchanged by the linear transformation of the matrix A.

1.2 Characteristic Polynomials and Characteristic Equations

A fundamental aspect of the eigenvalues of a matrix can be found by re-arranging Equation 1.7. Taking the $\lambda\vec{v}$ to the left-hand side of the equals sign, and introducing the appropriately sized identity matrix, gives the matrix-vector equation

$$(A - \lambda I)\vec{v} = \vec{0} \tag{1.8}$$

which means that an eigenvector is in the null space of a matrix that is closely related to A.

If we take matrix and vector norms of Equation 1.8, and recall that $\vec{v} \neq 0 \Rightarrow \|\vec{v}\| > 0$, then we know that

$$\begin{aligned} \|A - \lambda I\| \, \|\vec{v}\| &= \|\vec{0}\| \\ \Rightarrow \|A - \lambda I\| &= 0 \end{aligned} \tag{1.9}$$

If the norm of a matrix is zero then the matrix is singular; this implies that its determinant is zero. The *characteristic polynomial* of a matrix $A_{[n\times n]}$ is

$$F(\lambda) \stackrel{\text{def}}{=} \det(A_{[n\times n]} - \lambda I_{[n\times n]})$$

It is easy to show, by mathematical induction, that this polynomial is of order n in the variable λ.

The *characteristic equation* of of a matrix $A_{[n\times n]}$ sets the characteristic polynomial to 0, so

$$F(\lambda) = \det[A_{[n\times n]} - \lambda I_{[n\times n]}] = 0 \tag{1.10}$$

Equation 1.10 is fundamental to understanding many of the properties of eigenvalues. Note that the characteristic equation does not mention the eigenvectors; these, which are conceptually related, are found by other processes.

The roots of the characteristic equation of a matrix are the eigenvalues of the matrix. This observation is useful for small matrices but, for large matrices, the roots of the characteristic equation are seldom found directly.

1.3 Eigenfacts: Matrix Properties and Eigenvalues

There are many special cases and relations for the properties of a matrix and its eigenvalues. Some, such as the matrix transpose and triangular shape, come simply from the determinant; others require considerable linear algebra to derive. For reference, supposing that a matrix $A_{[n\times n]}$ has eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, a few of the properties include:

Transpose: The eigenvalues of A^T are the eigenvalues of A

Inverse: If A is invertible, that is if it is nonsingular, then the eigenvalues of A^{-1} are $\{1/\lambda_1, 1/\lambda_2, \ldots, 1/\lambda_n\}$

Trace: The *trace* of A, which is the sum of the diagonal entries of A, is the sum of the eigenvalues of A, so

$$\mathrm{tr}(A) = \sum_{j=1}^{n} a_{jj} = \sum_{j=1}^{n} \lambda_j$$

Real Symmetric: If A is a symmetric matrix, with $A = A^T$, and every entry of A is real, then every eigenvalue of A is real

Real Asymmetric: If A is an asymmetric matrix, with $A \neq A^T$, and every entry of A is real, then all of the eigenvalues of A are either (a) real or (b) occur in complex conjugate pairs

Determinant: The determinant of A is the product of the the eigenvalues of A, so

$$\det[A] = \prod_{j=1}^{n} \lambda_j = (\lambda_1 \lambda_2 \cdots \lambda_n)$$

Triangular: If A is a triangular matrix, whether upper or lower, then its eigenvalues are the diagonal entries of A

Diagonal: A diagonal matrix is a symmetric triangular matrix, so both above properties are true

Similar: If $P_{[n\times n]}$ is a nonsingular matrix, then the eigenvalues of $A_{[n\times n]}$ are the eigenvalues of $C = P^{-1}AP$; from the above properties:

- the traces are equal, so $\mathrm{tr}(A) = \mathrm{tr}(C)$
- the determinants are equal, so $\det(A) = \det(C)$

Characteristic Coefficients: If the characteristic polynomial, using `MATLAB` notation, is

$$F(\lambda) = \lambda^n + c_2\lambda^{n-1} + \cdots + c_n\lambda^1 + c_{n+1}$$

then $c_2 = -\mathrm{tr}(A)$ and $c_{n+1} = (-1)^n \det(A)$

1.4 Eigenvectors

In this course we will use eigenvectors of a matrix as a key means of understanding how it transforms vectors. Equation 1.7 gives the basic definition of an eigenvector but is not directly applicable to calculation. One alternative might be to find an eigenvalue, for example as a root of the characteristic equation, and use its value to find the associated eigenvector from Equation 1.8.

This method would find some λ_j and then describe an eigenvector as in the *null space*, or *kernel*, of the matrix

$$[A - \lambda_j I]$$

There are known methods for finding such a nullspace but they are seldom used. A principal problem is avoiding the trivial solution of $\vec{v}_j = \vec{v}$. This can be done by hand, or by inspection, for sample 2×2 matrices but rapidly becomes unwieldy. One common method is to use the QR decomposition repeatedly, which finds an eigenvalue/eigenvector pair.

After the eigenvectors of a matrix have been found, they can be used to analyze both the matrix and the linear transformation that it represents.

Extra Notes

1.5 Extra Notes: Extensions to Linear Algebra

The usual operators of linear algebra include addition, multiplication, trace, determinant, and usually norms and transposition. There are many other operators too. We will sometimes use extensions that are defined here.

Definition: vectorization operator

For any matrix $A_{m\times n}$ that is partitioned column-wise into vectors as

$$A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{bmatrix}$$

the *vectorization operator* $\mathrm{vec}(A)$ is defined as

$$\mathrm{vec}(A) \stackrel{\mathrm{def}}{=} \begin{bmatrix} \vec{a}_1 \\ \vec{a}_2 \\ \vdots \\ \vec{a}_n \end{bmatrix} \tag{1.11}$$

Observation: The $\mathrm{vec}(\cdot)$ operator "unravels" a matrix in a column-major order. A `MATLAB` expression for this is simply

```
A(:)
```

Definition: diagonal operator of a matrix

For any square matrix $A_{n\times n}$, the *diagonal operator* $\mathrm{diag}(A)$ is defined as

$$\mathrm{diag}(A) \stackrel{\text{def}}{=} \begin{bmatrix} a_{11} \\ a_{22} \\ \vdots \\ a_{nn} \end{bmatrix} \tag{1.12}$$

Observation: The $\mathrm{diag}(\cdot)$ operator of a matrix "extracts" the diagonal entries into a vector. A `MATLAB` expression for this is

```
diag(A)
```

Definition: diagonal operator of a vector

For any vector $\vec{a} \in \mathbb{R}^n$, the *diagonal operator* $\mathrm{diag}(\vec{a})$ is defined as

$$\mathrm{diag}(\vec{a}) \stackrel{\text{def}}{=} \begin{bmatrix} a_1 & 0 & 0 & \cdots & 0 \\ 0 & a_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_n \end{bmatrix} \tag{1.13}$$

Observation: The $\mathrm{diag}(\cdot)$ operator of a vector creates a diagonal matrix from the entries of the vector. A `MATLAB` expression for this is

```
diag(a)
```

It is straightforward to verify, for a vector $\vec{a}$, that $\mathrm{diag}(\mathrm{diag}(\vec{a})) = \vec{a}$. Likewise, for a diagonal matrix D, we have $\mathrm{diag}(\mathrm{diag}(D)) = D$.

1.6 Extra Notes: Characterization of a Real Symmetric Matrix

Any real symmetric matrix B can be characterized as one of five kinds. Each characterization relates the eigenvalues to the quadratic form.

Assuming that $B = B^T$ is real, and that $\vec{u} \in \mathbb{R}^n$ is a non-zero vector $\vec{u} \neq \vec{0}$, these symbols and quadratic forms are equivalent:

Symbol	**Name**	**Eigenvalues**	**Quadratic Form**
$B \succ 0$	Positive definite	$\forall_j \lambda_j > 0$	$\vec{u}^T B \vec{u} > 0$
$B \succeq 0$	Positive semidefinite	$\forall_j \lambda_j \geq 0$	$\vec{u}^T B \vec{u} \geq 0$
$B \prec 0$	Negative definite	$\forall_j \lambda_j < 0$	$\vec{u}^T B \vec{u} < 0$
$B \preceq 0$	Negative semidefinite	$\forall_j \lambda_j \leq 0$	$\vec{u}^T B \vec{u} \leq 0$
	Indefinite	$(\exists_i \lambda_i > 0) \wedge (\exists_j \lambda_j < 0)$	

End of Extra Notes