MATH2210 Notebook 4 — Spring 2018

prepared by Professor Jenny Baglivo

4 MATH2210 Notebook 4

This notebook is concerned with further matrix concepts and their applications. In particular, we will study eigenvalues, eigenvectors, orthogonality and least squares. The notes correspond to material in Chapters 5 and 6 of the Lay textbook.

4.1 Eigenvalues and Eigenvectors

An eigenvalue is an "exceptional value" and an eigenvector is an "exceptional vector."

The prefix "eigen" comes from the German language meaning "owned by" or "peculiar to."

Applications of eigenvalues and eigenvectors first appeared in the literature in the 18^{th} century, although the prefix "eigen" was not used until the early part of the 20^{th} century by the mathematician David Hilbert.

4.1.1 Definitions and Geometric Interpretations

Let A be a square matrix of order n, and λ ("lambda") be a scalar.

1. *Eigenvalue of a Square Matrix:*

   The scalar λ is said to be an eigenvalue of A if

   $$A\boldsymbol{x} = \lambda\boldsymbol{x} \text{ for some nonzero vector } \boldsymbol{x}.$$

2. *Eigenvector of a Square Matrix:*

   If $\boldsymbol{x} \neq \boldsymbol{O}$ satisfies $A\boldsymbol{x} = \lambda\boldsymbol{x}$, then $\boldsymbol{x}$ is an *eigenvector* of A with eigenvalue λ.

Note on Eigenvectors and Spans:

If $\boldsymbol{x}$ is a nonzero eigenvector, then so is $c\boldsymbol{x}$ for each nonzero $c \in \mathbf{R}$ since

$$A(c\boldsymbol{x}) = c(A\boldsymbol{x}) = c(\lambda\boldsymbol{x}) = \lambda(c\boldsymbol{x}).$$

Thus, the nonzero vectors in Span$\{\boldsymbol{x}\}$ are all eigenvectors of A with eigenvalue λ.

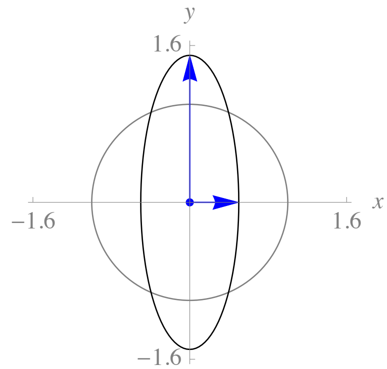
Note on Geometric Interpretation:

If $\boldsymbol{x} \in \mathbf{R}^n$ is eigenvector of A with eigenvalue $\lambda \in \mathbf{R}$, and if $T : \mathbf{R}^n \to \mathbf{R}^n$ is the linear transformation whose standard matrix is A, then T maps the span of $\boldsymbol{x}$ into itself:

$$T\left(\text{Span}\{\boldsymbol{x}\}\right) \subseteq \text{Span}\{\boldsymbol{x}\}.$$

Further, each vector is re-scaled by the eigenvalue λ.

Example 1. Let $A = \begin{bmatrix} 0.5 & 0 \\ 0 & 1.5 \end{bmatrix}$. Then



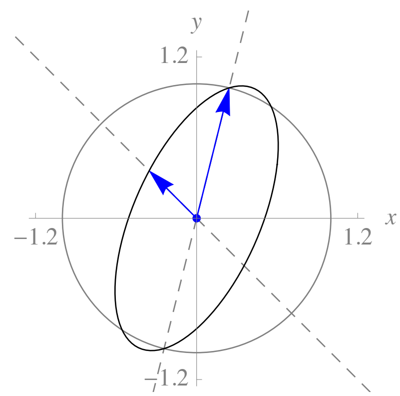
1. $\lambda_1 = 0.5$ is an eigenvalue of A, with corresponding eigenvector $\boldsymbol{e_1}$.
2. $\lambda_2 = 1.5$ is an eigenvalue of A, with corresponding eigenvector $\boldsymbol{e_2}$.

Consider the transformation with rule $T(\boldsymbol{x}) = A\boldsymbol{x}$.

Then

- $T(\boldsymbol{x})$ contracts points in the x-direction; in particular, $T(\boldsymbol{e_1}) = 0.5\boldsymbol{e_1}$.
- $T(\boldsymbol{x})$ expands points in the y-direction; in particular, $T(\boldsymbol{e_2}) = 1.5\boldsymbol{e_2}$.
- $T(\boldsymbol{x})$ maps the unit circle to the ellipse shown in the plot.

Example 2. Let $A = \begin{bmatrix} 0.6 & 0.1 \\ 0.4 & 0.9 \end{bmatrix}$. Then



1. $\lambda_1 = 0.5$ is an eigenvalue of A, with corresponding eigenvector $\boldsymbol{v_1} = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$,
2. $\lambda_2 = 1$ is an eigenvalue of A, with corresponding eigenvector $\boldsymbol{v_2} = \begin{bmatrix} 1/\sqrt{17} \\ 4/\sqrt{17} \end{bmatrix}$.

Consider the transformation with rule $T(\boldsymbol{x}) = A\boldsymbol{x}$.

Then

- $T(\boldsymbol{x})$ contracts points in the $\boldsymbol{v_1}$-direction; in particular, $T(\boldsymbol{v_1}) = 0.5\boldsymbol{v_1}$.
- $T(\boldsymbol{x})$ leaves points in the $\boldsymbol{v_2}$ direction fixed; in particular, $T(\boldsymbol{v_2}) = \boldsymbol{v_2}$.
- $T(\boldsymbol{x})$ maps the unit circle to the ellipse shown in the plot.

The complete analysis of Example 2 will be carried out in the next section.

Note on Choice of Examples:

These examples were chosen to emphasize that there is an attractive geometric interpretation when the matrix A has real eigenvalues ($\lambda \in \mathbf{R}$) and real eigenvectors ($\boldsymbol{x} \in \mathbf{R}^n$).

Some matrices, including those representing rotations around the origin through angles that are not integer multiples of π, leave no direction in 2-space fixed.

4.1.2 Eigenspaces, Characteristic Polynomials, Characteristic Equations

Let A be a square matrix of order n, and let λ be a scalar.

1. *Eigenspace of the Eigenvalue* λ*:*

   If λ is an eigenvalue of A, then the eigenspace of λ is defined as follows:

   $$\text{Eigenspace}(\lambda) = \{\boldsymbol{x} : A\boldsymbol{x} = \lambda\boldsymbol{x}\}.$$

   The eigenspace of λ contains all eigenvectors with eigenvalue λ and the zero vector. Since

   $$A\boldsymbol{x} = \lambda\boldsymbol{x} = \lambda I_n\boldsymbol{x} \quad \Leftrightarrow \quad A\boldsymbol{x} - \lambda I_n\boldsymbol{x} = \boldsymbol{O} \quad \Leftrightarrow \quad (A - \lambda I_n)\boldsymbol{x} = \boldsymbol{O},$$

   we know that Eigenspace(λ) $= Null(A - \lambda I_n)$. Further, since the eigenspace of λ must have positive dimension, we know that the matrix $(A - \lambda I_n)$ must be singular.

2. *Characteristic Polynomial of the Matrix* A*:*

   The characteristic polynomial of A is the expression $det(A - \lambda I)$. The characteristic polynomial is an n^{th} degree polynomial in the variable λ.

3. *Characteristic Equation of the Matrix* A*:*

   The characteristic equation of A is the equation $det(A - \lambda I) = 0$.

   To find the eigenvalues of A we solve the characteristic equation for λ.

Example 2, continued. Let $A = \begin{bmatrix} 0.6 & 0.1 \\ 0.4 & 0.9 \end{bmatrix}$, as above.

(1) To find the eigenvalues of A, we need to solve the characteristic equation.

Since

$$A - \lambda I = \begin{bmatrix} 0.6 & 0.1 \\ 0.4 & 0.9 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.6 - \lambda & 0.1 \\ 0.4 & 0.9 - \lambda \end{bmatrix},$$

the characteristic polynomial is

$$det(A - \lambda I) = (0.6 - \lambda)(0.9 - \lambda) - (0.1)(0.4) = 0.54 - 1.5\lambda + \lambda^2 - 0.04 = \lambda^2 - 1.5\lambda + 0.5.$$

Since the determinant can be factored,

$$det(A - \lambda I) = (\lambda - 0.5)(\lambda - 1),$$

the solutions to the characteristic equation are the eigenvalues 0.5 and 1.

(2) Let $\lambda = 0.5$.

Our goal is to solve $(A - 0.5I)\boldsymbol{x} = \boldsymbol{O}$.

Since

$$\left[\begin{array}{c|c} A - 0.5I & \boldsymbol{O} \end{array}\right] = \left[\begin{array}{cc|c} 0.1 & 0.1 & 0 \\ 0.4 & 0.4 & 0 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right] \Rightarrow \boldsymbol{x} = \begin{bmatrix} -x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

where x_2 is free, we know that

$$\text{Eigenspace}(0.5) = Null(A - 0.5I) = \text{Span}\left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}.$$

Note that I used

$$\boldsymbol{v_1} = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \in \text{Eigenspace}(0.5)$$

in the plot in the example.

(3) Let $\lambda = 1$.

Our goal is to solve $(A - I)\boldsymbol{x} = \boldsymbol{O}$.

Since

$$\left[\begin{array}{c|c} A - I & \boldsymbol{O} \end{array}\right] = \left[\begin{array}{cc|c} -0.4 & 0.1 & 0 \\ 0.4 & -0.1 & 0 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & -1/4 & 0 \\ 0 & 0 & 0 \end{array}\right] \Rightarrow \boldsymbol{x} = \begin{bmatrix} x_2/4 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 1/4 \\ 1 \end{bmatrix},$$

where x_2 is free, we know that

$$\text{Eigenspace}(1) = Null(A - I) = \text{Span}\left\{ \begin{bmatrix} 1/4 \\ 1 \end{bmatrix} \right\} = \text{Span}\left\{ \begin{bmatrix} 1 \\ 4 \end{bmatrix} \right\}.$$

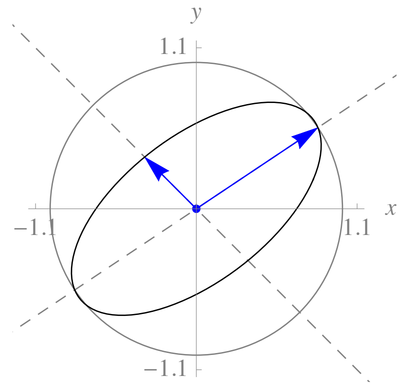
Note that I used

$$\boldsymbol{v_2} = \begin{bmatrix} 1/\sqrt{17} \\ 4/\sqrt{17} \end{bmatrix} = \frac{1}{\sqrt{17}} \begin{bmatrix} 1 \\ 4 \end{bmatrix} \in \text{Eigenspace}(1)$$

in the plot in the example.

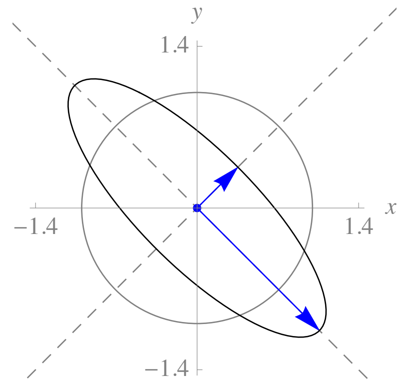
Problem 1. Let $A = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}$.

Determine the eigenvalues of A, and write each eigenspace as the span of a set of vectors.



Problem 2. Let $A = \begin{bmatrix} -0.5 & 1.0 \\ 1.0 & -0.5 \end{bmatrix}$.

Determine the eigenvalues of A, and write each eigenspace as the span of a set of vectors.



Problem 3. Let $A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0.8 & 0 \\ 0.25 & 0 & 0 \end{bmatrix}$.

Determine the eigenvalues of A, and write each eigenspace as the span of a set of vectors.

4.1.3 Eigenanalysis and Powers; Eigenvector Bases; Special Cases

Conducting an *eigenanalysis* (that is, finding eigenvalues and eigenvectors) can be challenging. The following is an initial list of useful theorems for eigenanalysis:

1. *Powers:* If λ is an eigenvalue of A with eigenvector $\boldsymbol{x}$ and k is a positive integer, then

   $$\boldsymbol{x} \text{ is an eigenvector for } A^k \text{ with corresponding eigenvalue } \lambda^k.$$

2. *Bases and Powers:* Let A be a square matrix of order n. Suppose that

   $$\boldsymbol{v_i} \text{ is an eigenvector for } A \text{ with corresponding eigenvalue } \lambda_i,$$

   for $i = 1, 2, \ldots, n$, and that the set $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}\}$ is a basis for $\mathbf{R}^n$.

   Then, for every vector $\boldsymbol{x}$ and positive integer k, $A^k\boldsymbol{x}$ can be computed quickly using the unique representation of $\boldsymbol{x}$ in the eigenvector basis $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}\}$.

   Specifically, if $\boldsymbol{x} = c_1\boldsymbol{v_1} + \cdots + c_n\boldsymbol{v_n}$ for unique constants c_i, then

   $$\begin{aligned} A^k\boldsymbol{x} &= A^k(c_1\boldsymbol{v_1} + \cdots + c_n\boldsymbol{v_n}) \\ &= c_1(A^k\boldsymbol{v_1}) + \cdots + c_n(A^k\boldsymbol{v_n}) \\ &= c_1(\lambda_1)^k\boldsymbol{v_1} + \cdots + c_n(\lambda_n)^k\boldsymbol{v_n}. \end{aligned}$$

3. *Diagonal Matrices:* Let A be a diagonal matrix of order n. Then

   $$\boldsymbol{e_i} \text{ is an eigenvector for } A \text{ with eigenvalue } a_{ii}, \text{ for } i = 1, 2, \ldots, n.$$

   Thus, the standard basis for $\mathbf{R}^n$ is an eigenvector basis for the diagonal matrix A, and the eigenvalues are the diagonal elements of A.

4. *Distinct Eigenvalues:* Let A be a square matrix of order n. If A has n distinct eigenvalues, then A has an eigenvector basis. To construct an eigenvector basis, choose one nonzero vector from each eigenspace.

5. *Triangular Matrices:* Let A be a triangular matrix of order n. Then

   $$det(A - \lambda I) = (a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda)$$

   and the eigenvalues of A are a_{11}, a_{22}, …, a_{nn}. (There may be repeats in the list.)

General application: projections over time. A general application of eigenanalysis is to the analysis of projections over time. In this type of application,

1. $\boldsymbol{x_0}$ represents information at time 0,
2. $\boldsymbol{x_1} = A\boldsymbol{x_0}$ represents information at time 1,
3. $\boldsymbol{x_2} = A\boldsymbol{x_1} = A^2\boldsymbol{x_0}$ represents information at time 2,

and so forth. If A has an eigenvector basis, then information at time k is

$$\boldsymbol{x_k} = A^k\boldsymbol{x_0} = c_1(\lambda_1)^k\boldsymbol{v_1} + \cdots + c_n(\lambda_n)^k\boldsymbol{v_n} \text{ where } \boldsymbol{x_0} = c_1\boldsymbol{v_1} + \cdots + c_n\boldsymbol{v_n}.$$

We will see important applications of this methodology later.

As a simple illustration, consider $A = \begin{bmatrix} 0.6 & 0.1 \\ 0.4 & 0.9 \end{bmatrix}$ once again. Let

$$\boldsymbol{v_1} = \begin{bmatrix} -1 \\ 1 \end{bmatrix},\ \boldsymbol{v_2} = \begin{bmatrix} 1 \\ 4 \end{bmatrix},\ \lambda_1 = 0.5,\ \lambda_2 = 1,\ \text{and } \boldsymbol{x_0} = c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2}.$$

Now,

$$\boldsymbol{x_k} = A^k\boldsymbol{x_0} = c_1(0.5)^k\boldsymbol{v_1} + c_2(1)^k\boldsymbol{v_2} \to c_2\boldsymbol{v_2} \text{ as } k \to \infty.$$

Thus, information at time k is approximately equal to the $\boldsymbol{v_2}$-component of information at time 0 when k is large.

Problem 4. Use the definitions of eigenvalue and eigenvector, and properties of matrices, to prove the following special case of the first theorem listed on the previous page:

> "Let A be a square matrix of order n, and let $\boldsymbol{x}$ be an eigenvector of A with eigenvalue λ. Demonstrate that $\boldsymbol{x}$ is an eigenvector of A^3 with eigenvalue λ^3."

Problem 5. Let A be a square matrix of order n with n distinct eigenvalues and let

$$\boldsymbol{v_i} \in \text{Eigenspace}(\lambda_i) \quad \text{be a nonzero vector, for each } i.$$

Demonstrate that $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}\}$ is a linearly independent set, thus forming a basis for $\mathbf{R}^n$.

Problem 6. Each triangular matrix below has eigenvalues 3, 3, 2 (including multiplicities).

In each case, write Eigenspace(3) and Eigenspace(2) as spans of sets of vectors.

$(a)\ A = \begin{bmatrix} 3 & 0 & 0 \\ 2 & 3 & 0 \\ 2 & 1 & 2 \end{bmatrix}$

$$(b)\ A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ -1 & 1 & 2 \end{bmatrix}$$

4.1.4 Algebraic Multiplicity and Geometric Multiplicity

Let A be a square matrix of order n and let λ_0 be an eigenvalue of A. Then

1. *Algebraic Multiplicity:*

   The algebraic multiplicity of λ_0 is the number of times $(\lambda_0 - \lambda)$ appears as a factor of the characteristic polynomial.

2. *Geometric Multiplicity:*

   The geometric multiplicity of λ_0 is the dimension of Eigenspace(λ_0).

Problem 6, continued. Fill-in the table below with information for the triangular matrices from the problem beginning on page 13:

	Geometric Multiplicity of $\lambda = 2$	Algebraic Multiplicity of $\lambda = 2$	Geometric Multiplicity of $\lambda = 3$	Algebraic Multiplicity of $\lambda = 3$
$(a)\quad A = \begin{bmatrix} 3 & 0 & 0 \\ 2 & 3 & 0 \\ 2 & 1 & 2 \end{bmatrix}$				
$(b)\quad A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ -1 & 1 & 2 \end{bmatrix}$				

Footnote on The Fundamental Theorem of Algebra:

The fundamental theorem of algebra implies that the characteristic polynomial of the matrix A can be factored into n linear terms if we allow both real and complex numbers:

$$det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda), \text{ where each } \lambda_i \in \boldsymbol{C}.$$

Thus, the fundamental theorem of algebra implies that the sum of the algebraic multiplicities of A must be n (the order of the matrix).

Matrices with complex eigenvalues and eigenvectors are common in applied mathematics. Examples include population projection matrices (see Section 4.1.7, page 19).

4.1.5 Multiplicity and Finding Eigenvector Bases

Here are two additional theorems that are useful for doing eigenanalyses:

1. *Algebraic and Geometric Multiplicity:* Let A be a square matrix of order n and let λ_0 be an eigenvalue of A. Then the algebraic and geometric multiplicities of λ_0 must satisfy the following inequalities:

   $$1 \leq \text{Geometric Multiplicity of } \lambda_0 \leq \text{Algebraic Multiplicity of } \lambda_0.$$

   (If the geometric multiplicity is strictly less than the algebraic multiplicity, then there is a "deficiency" of eigenvectors and we won't be able to find an eigenvector basis.)

2. *Pooling Eigenspace Bases:* If A has p distinct eigenvalues (λ_i for $i = 1, 2, \ldots, p$) and $\mathcal{B}_i$ is a basis for eigenspace of λ_i for each i, then the set

   $$\mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_p$$

   is a linearly independent set. (If you pool the bases, you get a linearly independent set.)

Problem 6, continued. Do either of the matrices in Problem 6 have an eigenvector basis for $\mathbf{R}^3$? If yes, explicitly write down a basis. If no, explain why.

4.1.6 Similar Matrices and Diagonalizable Matrices

Let A and B be square matrices of order n, and let D be a diagonal matrix of order n.

1. *Similar Matrices:*
   A and B are said to be similar if there exists an invertible matrix P satisfying
   $$A = PBP^{-1} \text{ (and } B = P^{-1}AP\text{)}.$$
2. *Diagonalizable Matrix:*
   A is said to be diagonalizable if it is similar to a diagonal matrix. That is, if
   $$A = PDP^{-1}, \text{ where } P \text{ is invertible,}$$
   for some diagonal matrix D.

If A and B are similar matrices, then

1. they have the same determinant, $det(A) = det(B)$,
2. they have the same eigenvalues, and
3. their k^{th} powers satisfy $A^k = PB^kP^{-1}$, for each positive integer k.

The factorization $A = PBP^{-1}$ is useful when B is easier to work with than A.

Question: Can you see why similar matrices have the same eigenvalues?

The following theorem tells us exactly when A is diagonalizable.

Diagonalization Theorem. Let A be a square matrix of order n. Then

A is diagonalizable if and only if A has n linearly independent eigenvectors.

In fact, $A = PDP^{-1}$ if and only if the columns of P are n linearly independent eigenvectors and the diagonal entries of D are the corresponding eigenvalues.

For example, using the work from page 9, if $A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0.8 & 0 \\ 0.25 & 0 & 0 \end{bmatrix}$, then

$$A = PDP^{-1} \quad \text{where} \quad P = \begin{bmatrix} 0 & 2 & -2 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 0.8 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & -0.5 \end{bmatrix}.$$

Relationship to Transformations: Suppose that $A = PDP^{-1}$, where

$$P = \begin{bmatrix} \boldsymbol{v_1} & \boldsymbol{v_2} & \dots & \boldsymbol{v_n} \end{bmatrix} \text{ and } D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix},$$

and let $\mathcal{B} = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n}\}$ be the basis whose elements are the columns of P.

Then the "action" of A is to

($\downarrow$): change from the standard basis of $\mathbf{R}^n$ to basis $\mathcal{B}$;

($\rightarrow$): operate as a diagonal matrix in basis $\mathcal{B}$, and

($\uparrow$): change back to the standard basis for interpretation.

$$\begin{array}{ccc} \boldsymbol{x} = \sum_{i=1}^n c_i \boldsymbol{v_i} & \xrightarrow{A} & A\boldsymbol{x} = \sum_{i=1}^n \lambda_i c_i \boldsymbol{v_i} \\ \downarrow & & \uparrow \\ [\boldsymbol{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} & \xrightarrow{D} & [A\boldsymbol{x}]_{\mathcal{B}} = \begin{bmatrix} \lambda_1 c_1 \\ \lambda_2 c_2 \\ \vdots \\ \lambda_n c_n \end{bmatrix} \end{array}$$

Similarly, the "action" of A^k is to

($\downarrow$): change from the standard basis of $\mathbf{R}^n$ to basis $\mathcal{B}$;

($\rightarrow$): operate as the k^{th} power of a diagonal matrix in basis $\mathcal{B}$, and

($\uparrow$): change back to the standard basis for interpretation.

$$\begin{array}{ccc} \boldsymbol{x} = \sum_{i=1}^n c_i \boldsymbol{v_i} & \xrightarrow{A^k} & A^k\boldsymbol{x} = \sum_{i=1}^n \lambda_i^k c_i \boldsymbol{v_i} \\ \downarrow & & \uparrow \\ [\boldsymbol{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} & \xrightarrow{D^k} & [A^k\boldsymbol{x}]_{\mathcal{B}} = \begin{bmatrix} \lambda_1^k c_1 \\ \lambda_2^k c_2 \\ \vdots \\ \lambda_n^k c_n \end{bmatrix} \end{array}$$

4.1.7 Applications: Population Projections and Stochastic Matrices

This section reviews concepts and introduces two applications of eigenanalysis.

I. Concepts needed for applications

The *Fundamental Theorem of Algebra* implies that we can factor the characteristic polynomial of a $n \times n$ matrix A into n linear terms if we allow both real and complex numbers:

$$det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda), \quad \text{where each } \lambda_i \in \boldsymbol{C}.$$

The eigenvalues of A are λ_1, λ_2, ..., λ_n. In general, not all λ_i's are distinct.

Exercise. Factor the characteristic polynomial of $A = \begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ -4 & 0 & 0 & 1 \end{bmatrix}$ into 4 linear terms.

List the eigenvalues of A, with multiplicities.

Absolute values and ordering eigenvalues. The eigenvalues of a square matrix are often ordered by their absolute values from largest to smallest:

$$|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \cdots |\lambda_n|,$$

where the absolute value of a complex number is defined as $\boxed{|a+bi| = \sqrt{a^2+b^2}}$.

Exercise, cntd. Since the eigenvalues of the 4×4 matrix A are __________________________,

we know that their absolute values are ____.

Applications of diagonalization. In each application, A can be written as $A = PDP^{-1}$, where

$$P = \begin{bmatrix} \boldsymbol{v_1} & \cdots \boldsymbol{v_n} \end{bmatrix} \quad \text{is a matrix whose columns form an eigenvector basis}$$

D is the diagonal matrix of corresponding eigenvalues, and the eigenvectors and eigenvalues have been ordered by decreasing absolute value. Further, in each application, the first eigenvalue is *strictly dominant*, that is,

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots |\lambda_n|.$$

We let $\boldsymbol{x_0} = c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_n\boldsymbol{v_n}$ represent information at time 0,

- $\boldsymbol{x_1} = A\boldsymbol{x_0}$ represent information at time 1,
- $\boldsymbol{x_2} = A\boldsymbol{x_1} = A^2\boldsymbol{x_0}$ represent information at time 2,
- $\boldsymbol{x_3} = A\boldsymbol{x_2} = A^2\boldsymbol{x_1} = A^3\boldsymbol{x_0}$ represent information at time 3,

and so forth. Since $A^k = PD^kP^{-1}$, we know that

$$\boldsymbol{x_k} = A^k\boldsymbol{x_0} = c_1(\lambda_1)^k\boldsymbol{v_1} + c_2(\lambda_2)^k\boldsymbol{v_2} + \cdots + c_n(\lambda_n)^k\boldsymbol{v_n}$$

and we look for patterns in the powers of the eigenvalues.

II. Population projections, and the northern spotted owl (***Source:*** **Lay text, p.267).**

Researchers used demographic data for the northern spotted owl to develop a stage-matrix model using 3 life stages (juvenile, subadult and adult). Their goal was to track the population growth/decline of the owl in a particular old growth forest in the Pacific northwest.

If j_i is the number of juveniles, s_i is the number of subadults and a_i is the number of adults in the population at time i, then

$$\boldsymbol{x_i} = \begin{bmatrix} j_i \\ s_i \\ a_i \end{bmatrix} \text{ is the population vector at time } i,$$

and the total population at time i is the sum of the components $(j_i + s_i + a_i)$. The matrix that allows you to project one year is

$$A = \begin{bmatrix} 0 & 0 & 0.33 \\ 0.18 & 0 & 0 \\ 0 & 0.71 & 0.94 \end{bmatrix}.$$

Given $\boldsymbol{x_i}$, the population vector at time $(i+1)$ is

$$\boldsymbol{x_{i+1}} = A\boldsymbol{x_i} = \begin{bmatrix} 0 & 0 & 0.33 \\ 0.18 & 0 & 0 \\ 0 & 0.71 & 0.94 \end{bmatrix} \begin{bmatrix} j_i \\ s_i \\ a_i \end{bmatrix} = \begin{bmatrix} 0.33a_i \\ 0.18j_i \\ 0.71s_i + 0.94a_i \end{bmatrix} = \begin{bmatrix} j_{i+1} \\ s_{i+1} \\ a_{i+1} \end{bmatrix}.$$

Matrices used in population problems are generally diagonalizable, with both real and complex eigenvalues. For this problem, we can write $A = PDP^{-1}$, where

$$D = \begin{bmatrix} 0.984 & 0 & 0 \\ 0 & -0.022 + 0.206i & 0 \\ 0 & 0 & -0.022 - 0.206i \end{bmatrix}$$

and

$$P = \begin{bmatrix} 0.318 & 0.682 & 0.682 \\ 0.058 & -0.062 - 0.59i & -0.062 + 0.59i \\ 0.946 & -0.045 + 0.426i & -0.045 - 0.426i \end{bmatrix}.$$

The diagonal elements of D are the eigenvalues of A. Since

$$(0.984)^k \to 0,\ (-0.022 + 0.206i)^k \to 0 \text{ and } (-0.022 - 0.206i)^k \to 0 \text{ as } k \to \infty,$$

we know that the population will eventually crash given any initial population vector.

The author tells us that, if the (2,1)-entry of the A matrix was 0.60 (the proportion that would be appropriate for this species in a different location), then the population would grow. The new matrix (the one with the new (2,1)-entry) would be diagonalizable with

$$D = \begin{bmatrix} 1.064 & 0 & 0 \\ 0 & -0.062 + 0.358i & 0 \\ 0 & 0 & -0.062 - 0.358i \end{bmatrix}$$

and

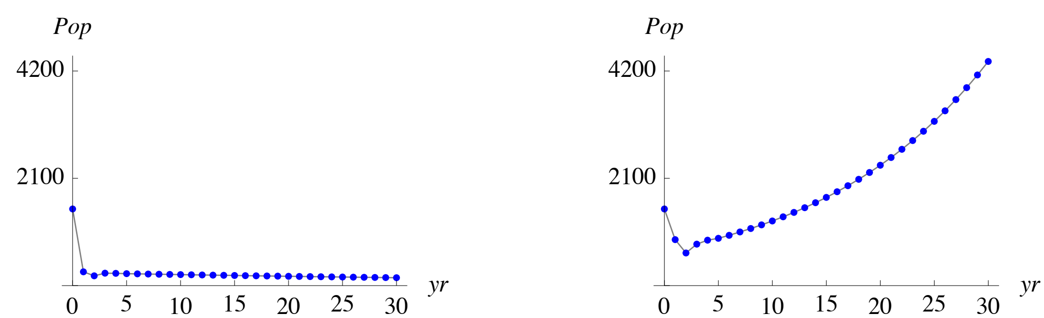
$$P = \begin{bmatrix} 0.292 & -0.077 + 0.443i & -0.077 - 0.443i \\ 0.165 & 0.743 & 0.743 \\ 0.942 & -0.467 - 0.167i & -0.467 + 0.167i \end{bmatrix}.$$

Since

$$(1.064)^k \to \infty, \ (-0.062 + 0.358i)^k \to 0 \text{ and } (-0.062 - 0.358i)^k \to 0 \text{ as } k \to \infty,$$

the statement the author makes is correct.

I simulated population growth over 30 years starting with an initial population of 1500 individuals ($j_0 + s_0 + a_0 = 1500$) and using both the true matrix and the matrix with the (2,1)-entry changed. The results are summarized in the plot below.



1. *Left Plot:* Using the true matrix, the total population declined over time.
2. *Right Plot:* Using the altered matrix, the total population increased over time with approximate geometric growth. Geometric growth "kicks in" when k is large enough so that k^{th} powers of the last two eigenvalues are close to zero:

$$(-0.062 + 0.358i)^k \approx 0 \text{ and } (-0.062 - 0.358i)^k \approx 0.$$

III. Stochastic matrices, moving cars, and searching the internet

A *probability vector* is one whose entries are nonnegative real numbers with sum 1. A *stochastic matrix* is a square matrix whose columns are probability vectors.

The following matrices are examples of stochastic matrices:

$$\begin{bmatrix} 0.9 & 0.3 \\ 0.1 & 0.7 \end{bmatrix} \qquad \begin{bmatrix} 0.25 & 0.2 & 0 \\ 0.5 & 0.5 & 1 \\ 0.25 & 0.3 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0 & 0.4 \end{bmatrix}.$$

Stochastic matrices are used to model population movement over time, where individuals move among n different locations.

The following simple example considers the movement of rental cars over time.

Example. A car rental agency has three rental locations (1, 2, 3). A customer may rent a car from any of the three locations and return the car to any of the three locations. From past experience, management observes that:

1. *Location 1:* Cars rented from location 1 are returned to locations 1, 2, 3 with probabilities 0.5, 0.3 and 0.2, respectively;
2. *Location 2:* Cars rented from location 2 are returned to locations 1, 2, 3 with probabilities 0.2, 0.8 and 0, respectively; and
3. *Location 3:* Cars rented from location 3 are returned to locations 1, 2, 3 with probabilities 0.3, 0.3 and 0.4, respectively.

Suppose that we would like to determine the probabilities that a car initially rented from a given location (either 1, 2 or 3) will be returned to locations 1, 2, 3 after k rental periods. Let A be the matrix whose columns are the probabilities listed above, let a_i, b_i and c_i be the probabilities that the car is at locations 1, 2, 3 after i rental periods, and let $\boldsymbol{x_i}$ be the vector whose components are the probabilities a_i, b_i and c_i:

$$A = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0 & 0.4 \end{bmatrix}, \quad \boldsymbol{x_i} = \begin{bmatrix} a_i \\ b_i \\ c_i \end{bmatrix}.$$

The matrix A can be used to project one rental period. That is, $\boldsymbol{x_{i+1}} = A\boldsymbol{x_i}$ for each i. The starting location vectors ($\boldsymbol{x_0}$) are

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ for location 1, } \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \text{ for location 2, } \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \text{ for location 3.}$$

We can write $A = PDP^{-1}$, where

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.2 \end{bmatrix}, \quad P = \begin{bmatrix} \boldsymbol{v_1} & \boldsymbol{v_2} & \boldsymbol{v_3} \end{bmatrix} = \begin{bmatrix} 0.3 & 0.1 & -0.1 \\ 0.6 & -0.3 & 0.0 \\ 0.1 & 0.2 & 0.1 \end{bmatrix},$$

and the first column of P has nonnegative terms with sum 1. Note that

$$1^k \to 1, \quad (0.5)^k \to 0 \ \text{ and } \ (0.2)^k \to 0 \ \text{ as } \ k \to \infty.$$

If $\boldsymbol{x_0}$ corresponds to the location 2 vector, for example, then

$$\boldsymbol{x_0} = \boldsymbol{v_1} - (4/3)\boldsymbol{v_2} + (5/3)\boldsymbol{v_3}$$

and $\boldsymbol{x_k} = A^k \boldsymbol{x_0} \approx \boldsymbol{v_1}$ for large k. In fact, $\boldsymbol{x_k} \approx \boldsymbol{v_1}$ after only 10 time periods:

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$	$k=11$
a_k	0	0.2	0.26	0.282	0.291	0.296	0.298	0.299	0.299	0.3	0.3	0.3
b_k	1	0.8	0.7	0.65	0.625	0.613	0.606	0.603	0.602	0.601	0.6	0.6
c_k	0	0	0.04	0.068	0.084	0.092	0.096	0.098	0.099	0.099	0.1	0.1

Similarly, if $\boldsymbol{x_0}$ corresponds to the location 1 vector, then

$$\boldsymbol{x_0} = \boldsymbol{v_1} + 2\boldsymbol{v_2} - 5\boldsymbol{v_3}$$

and $\boldsymbol{x_k} = A^k \boldsymbol{x_0} \approx \boldsymbol{v_1}$ for large k, and if $\boldsymbol{x_0}$ corresponds to the location 3 vector, then

$$\boldsymbol{x_0} = \boldsymbol{v_1} + 2\boldsymbol{v_2} + 5\boldsymbol{v_3}$$

and $\boldsymbol{x_k} = A^k \boldsymbol{x_0} \approx \boldsymbol{v_1}$ for large k. Thus, if k is large, the probabilities that a car initially at any one of the three locations will be returned to locations 1, 2, 3 after k rental periods are (approximately) 0.3, 0.6 and 0.1.

Surfing the web. Now, imagine yourself surfing the web starting from some initial location and randomly following hyperlinks. Assuming an appropriate A matrix can be created and analyzed as above, the probability that you will be at a given location after a sufficient number of steps can be determined.

The designers of *Google* use the eventual probabilities to determine the order in which the results of a search are reported; specifically, webpages with higher probabilities are listed before those with lower probabilities. Their A matrix uses the hyperlink structure of the web and some proprietary information.

4.2 Orthogonality and Orthogonal Projections

4.2.1 Inner Product, Length and Distance

Let $\boldsymbol{v}$ and $\boldsymbol{w}$ be vectors in $\mathbf{R}^k$. Then

1. *Inner Product of Two Vectors:*

   The inner product (or *dot product*) of $\boldsymbol{v}$ and $\boldsymbol{w}$ is the number

$$\boldsymbol{v}\cdot\boldsymbol{w} = \boldsymbol{v}^{\boldsymbol{T}}\boldsymbol{w} = \begin{bmatrix} v_1 & v_2 & \cdots & v_k \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix} = v_1w_1 + v_2w_2 + \cdots + v_kw_k.$$

2. *Length of A Vector:*

   The length of $\boldsymbol{v}$ is the number

$$\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v}\cdot\boldsymbol{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_k^2}.$$

3. *Distance Between Two Vectors:*

   The distance between $\boldsymbol{v}$ and $\boldsymbol{w}$ is the length of the difference vector $\boldsymbol{v} - \boldsymbol{w}$:

$$dist(\boldsymbol{v}, \boldsymbol{w}) = \|\boldsymbol{v} - \boldsymbol{w}\| = \sqrt{(v_1 - w_1)^2 + (v_2 - w_2)^2 + \cdots + (v_k - w_k)^2}.$$

4. *Unit Vector in the Direction of* $\boldsymbol{v}$*:*

   If $\boldsymbol{v} \neq \boldsymbol{O}$, then $\boldsymbol{u} = \dfrac{1}{\|\boldsymbol{v}\|}\boldsymbol{v}$ is the unit vector in the direction of $\boldsymbol{v}$.

   Note that the length of $\boldsymbol{u}$ is one.

For example, if $\boldsymbol{v} = \begin{bmatrix} 8 \\ -1 \\ 0 \\ 2 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} -2 \\ -1 \\ 4 \\ 3 \end{bmatrix}$, then

(1) $\boldsymbol{v}\cdot\boldsymbol{w} =$ ________________________________

(2) The length of $\boldsymbol{v}$ is ________________________________

(3) The unit vector in the direction of $\boldsymbol{v}$ is ________________________________

(4) The distance between $\boldsymbol{v}$ and $\boldsymbol{w}$ is ________________________________

4.2.2 Properties of Inner Product

Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbf{R}^k$, $c \in \mathbf{R}$. Then

1. *Commutative:* $\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{w} \cdot \boldsymbol{v}$.
2. *Scalars:* $(c\boldsymbol{v}) \cdot \boldsymbol{w} = \boldsymbol{v} \cdot (c\boldsymbol{w}) = c(\boldsymbol{v} \cdot \boldsymbol{w})$.
3. *Distributive:* $(\boldsymbol{u} + \boldsymbol{v}) \cdot \boldsymbol{w} = (\boldsymbol{u} \cdot \boldsymbol{w}) + (\boldsymbol{v} \cdot \boldsymbol{w})$ and $\boldsymbol{w} \cdot (\boldsymbol{u} + \boldsymbol{v}) = (\boldsymbol{w} \cdot \boldsymbol{u}) + (\boldsymbol{w} \cdot \boldsymbol{v})$.
4. *Nonnegative:* $\boldsymbol{v} \cdot \boldsymbol{v} \geq 0$. Further $\boldsymbol{v} \cdot \boldsymbol{v} = 0$ iff $\boldsymbol{v} = \boldsymbol{O}$.

Problem 1. Let $\boldsymbol{v_1}$, $\boldsymbol{v_2}$ and $\boldsymbol{w}$ be vectors in $\mathbf{R}^k$, and suppose that $\boldsymbol{v_1} \cdot \boldsymbol{w} = 0$ and $\boldsymbol{v_2} \cdot \boldsymbol{w} = 0$. Use the properties of inner product to demonstrate that $\boldsymbol{v} \cdot \boldsymbol{w} = 0$ for every $\boldsymbol{v} \in \text{Span}\,\{\boldsymbol{v_1}, \boldsymbol{v_2}\}$.

4.2.3 Orthogonal Vectors, Orthogonal Sets and Orthogonal Complement

The concept of orthogonality is important in applications.

1. *Orthogonal Vectors:*

   The vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbf{R}^k$ are said to be orthogonal if $\boldsymbol{v} \cdot \boldsymbol{w} = 0$.

2. *Orthogonal Set of Vectors:*

   The set $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_p}\} \subset \mathbf{R}^k$ is said to be an orthogonal set if

   $$\boldsymbol{v_i} \neq \boldsymbol{O} \text{ for all } i, \text{ and } \boldsymbol{v_i} \cdot \boldsymbol{v_j} = 0 \text{ when } i \neq j.$$

3. *Orthogonal Complement of a Subspace:*

   Let V be a subspace of $\mathbf{R}^k$. The orthogonal complement of V, denoted by $V^\perp$ ("V-perp"), is the collection of all vectors orthogonal to V:

   $$V^\perp = \{\boldsymbol{w} \ : \ \boldsymbol{w} \text{ is orthogonal to each } \boldsymbol{v} \in V\}.$$

   Both V and $V^\perp$ are subsets of $\mathbf{R}^k$.

Problem 2. Let $\boldsymbol{v_1} = \begin{bmatrix} a \\ -2 \\ 1 \end{bmatrix}$, $\boldsymbol{v_2} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$, $\boldsymbol{v_3} = \begin{bmatrix} 5 \\ 2 \\ b \end{bmatrix}$.

Find values of a, b so that $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}$ is an orthogonal set.

Properties of orthogonal complements. Let V be a subspace of $\mathbf{R}^k$ and let $V^\perp$ be its orthogonal complement. Then

1. *Subspace:* The orthogonal complement of V is a subspace of $\mathbf{R}^k$. Further,

$$V \cap V^\perp = \{\boldsymbol{O}\} \text{ since } \boldsymbol{O} \text{ is the only vector satisfying } \boldsymbol{x}\cdot\boldsymbol{x} = 0.$$

2. *Orthogonal Complement of $V^\perp$:* The orthogonal complement of $V^\perp$ is V: $\left(V^\perp\right)^\perp = V$.

3. *Spanning Sets:* Suppose that $V = \text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_p}\}$. Then

$$\boldsymbol{w} \in V^\perp \text{ if and only if } \boldsymbol{w}\cdot\boldsymbol{v_i} = 0 \text{ for } i = 1, 2, \ldots, p.$$

4. *Pooling Bases:* If $\mathcal{B}_1$ is a basis for V and $\mathcal{B}_2$ is a basis for $V^\perp$, then the union of the bases, $\mathcal{B}_1 \cup \mathcal{B}_2$, is a basis for $\mathbf{R}^k$.

To illustrate orthogonal complements, let

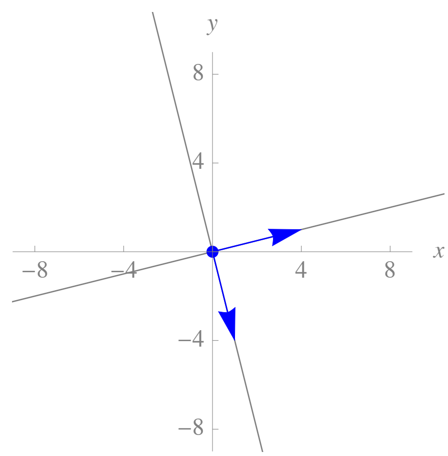
$$V = \text{Span}\{\boldsymbol{v}\} = \text{Span}\left\{\begin{bmatrix}4\\1\end{bmatrix}\right\},$$

and $\boldsymbol{w} = \begin{bmatrix}w_1\\w_2\end{bmatrix}$. Since

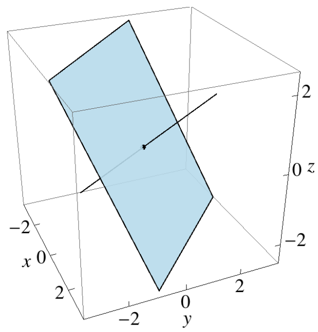
$$0 = \boldsymbol{w}\cdot\boldsymbol{v} = 4w_1 + w_2 \;\Rightarrow\; \boldsymbol{w} = w_1\begin{bmatrix}1\\-4\end{bmatrix},$$

where w_1 is free, we know that

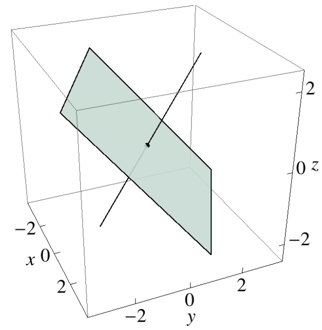
$$V^\perp = \text{Span}\left\{\begin{bmatrix}1\\-4\end{bmatrix}\right\}, \text{ as shown in the plot.}$$



Problem 3. In each case, write $V^{\perp}$ as a span.



$$(a)\quad V = \mathrm{Span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} \right\}$$



$$(b)\quad V = \mathrm{Span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\}$$

4.2.4 Fundamental Theorem of Linear Algebra

Let A be an $m \times n$ matrix and let A^T be its transpose. The following theorem gives important relationships among the four subspaces related to A and its transpose.

Fundamental Theorem of Linear Algebra. Let A be an $m \times n$ matrix. Then

1. $Null(A)$ and $Col(A^T)$ are orthogonal complements in $\mathbf{R}^n$.
2. $Null(A^T)$ and $Col(A)$ are orthogonal complements in $\mathbf{R}^m$.

Further, if $rank(A) = r$, then

- $dim(Col(A)) = dim(Col(A^T)) = r$,
- $dim(Null(A)) = n - r$ and
- $dim(Null(A^T)) = m - r$.

It is instructive to demonstrate that $Null(A)$ and $Col(A^T)$ are orthogonal complements in $\mathbf{R}^n$.

Let $A^T = \begin{bmatrix} \boldsymbol{\alpha_1} & \boldsymbol{\alpha_2} & \cdots & \boldsymbol{\alpha_m} \end{bmatrix}$ and $A = \begin{bmatrix} \boldsymbol{\alpha_1}^T \\ \vdots \\ \boldsymbol{\alpha_m}^T \end{bmatrix}$. Then

$$A\boldsymbol{x} = \boldsymbol{O} \quad \Leftrightarrow \quad \begin{bmatrix} \boldsymbol{\alpha_1}^T \\ \boldsymbol{\alpha_2}^T \\ \vdots \\ \boldsymbol{\alpha_m}^T \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} \boldsymbol{\alpha_1} \cdot \boldsymbol{x} \\ \boldsymbol{\alpha_2} \cdot \boldsymbol{x} \\ \vdots \\ \boldsymbol{\alpha_m} \cdot \boldsymbol{x} \end{bmatrix} = \boldsymbol{O}.$$

Now (complete the proof),

Problem 4. Find bases for $Null(A)$, $Col(A)$, $Null(A^T)$ and $Col(A^T)$, where

$$A = \begin{bmatrix} 1 & 2 & 1 & -2 \\ -3 & -6 & -4 & 3 \\ 2 & 4 & 1 & -7 \end{bmatrix}.$$

4.2.5 Orthogonal Spanning Sets

The following theorem tells us that a set of mutually orthogonal nonzero vectors is linearly independent. Further, the coordinates of a vector $\boldsymbol{w}$ with respect to a basis of mutually orthogonal nonzero vectors can be found quickly using dot products:

> ***Theorem (Orthogonal Spanning Sets).*** Let $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_p}\}$ be an orthogonal set of vectors in $\mathbf{R}^k$ and let $V = \text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_p}\}$. Then
>
> 1. $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_p}\}$ is a basis for V.
> 2. If $\boldsymbol{w} \in V$, then $\boldsymbol{w} = c_1\boldsymbol{v_1} + \cdots + c_p\boldsymbol{v_p}$ where $c_i = \frac{\boldsymbol{w}\cdot\boldsymbol{v_i}}{\boldsymbol{v_i}\cdot\boldsymbol{v_i}}$ for each i.

Problem 5. In each case, use dot products to find the coordinates of the vector $\boldsymbol{w}$ with respect to the given orthogonal basis.

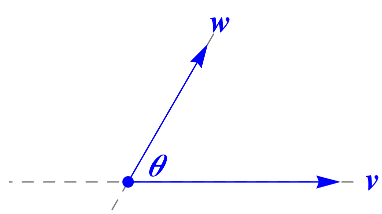
(a) $V = \text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}\} = \text{Span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} \right\}$, and $\boldsymbol{w} = \begin{bmatrix} 14 \\ 1 \\ 11 \end{bmatrix}$.

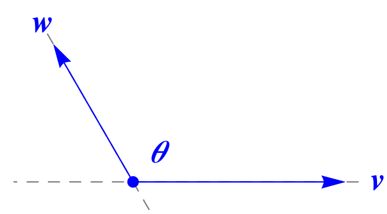
(*b*) $V = \text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\} = \text{Span}\left\{\begin{bmatrix} 0 \\ 1 \\ -4 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ -4 \end{bmatrix}\right\}$, and $\boldsymbol{w} = \begin{bmatrix} 7 \\ 13 \\ 0 \\ 4 \end{bmatrix}$.

4.2.6 Angles, Inner Products, and Orthogonal Projections

Angle between $\boldsymbol{v}$ and $\boldsymbol{w}$. Let $\boldsymbol{v}$ and $\boldsymbol{w}$ be vectors in $\mathbf{R}^2$ or $\mathbf{R}^3$ represented as directed line segments beginning at the origin. The *angle* between $\boldsymbol{v}$ and $\boldsymbol{w}$, θ, is the smaller of the two angles at the origin determined by $\boldsymbol{v}$ and $\boldsymbol{w}$. The angle θ lies in the interval $[0, \pi]$.

The following plots illustrate angles satisfying $0 < \theta < \frac{\pi}{2}$ (*left*) and $\frac{\pi}{2} < \theta < \pi$ (*right*):





Analytic geometry can be used to demonstrate that $\boxed{\boldsymbol{v}\cdot\boldsymbol{w} = \|\boldsymbol{v}\|\ \|\boldsymbol{w}\|\ \cos(\theta)}$.

Note, in particular, that if $\theta = \frac{\pi}{2}$, then $\boldsymbol{v}\cdot\boldsymbol{w} = 0$.

Orthogonal projection. For vectors in $\mathbf{R}^2$ or $\mathbf{R}^3$, the *(orthogonal) projection* of $\boldsymbol{w}$ onto $\boldsymbol{v}$, denoted by $\text{proj}_{\boldsymbol{v}}(\boldsymbol{w})$, is the vector highlighted in each diagram below for angles $\theta \neq \frac{\pi}{2}$:

1. *Left Plot* $\left(0 \leq \theta < \frac{\pi}{2}\right)$*:* The projection of $\boldsymbol{w}$ onto $\boldsymbol{v}$ is the vector that points in the direction of $\boldsymbol{v}$ and whose length is $\|\boldsymbol{w}\| \cos(\theta)$.
2. *Right Plot* $\left(\frac{\pi}{2} < \theta \leq \pi\right)$*:* The projection of $\boldsymbol{w}$ onto $\boldsymbol{v}$ is the vector that points in the direction opposite to $\boldsymbol{v}$ and whose length is $\|\boldsymbol{w}\| \cos(\pi - \theta)$.

When $\theta = \frac{\pi}{2}$, the projection of $\boldsymbol{w}$ onto $\boldsymbol{v}$ is the zero vector.

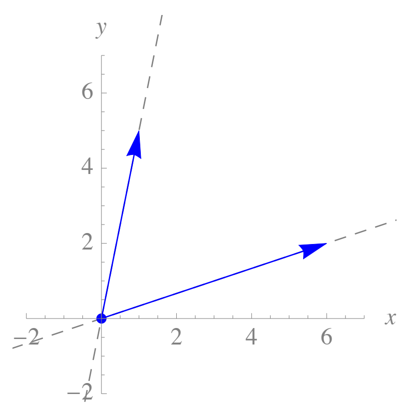
Geometry, trigonometry and the relationship $\boldsymbol{v} \cdot \boldsymbol{w} = \|\boldsymbol{v}\| \, \|\boldsymbol{w}\| \cos(\theta)$ can be used to demonstrate that the projection can be computed as follows:

$$\text{proj}_{\boldsymbol{v}}(\boldsymbol{w}) = \left(\frac{\boldsymbol{w} \cdot \boldsymbol{v}}{\boldsymbol{v} \cdot \boldsymbol{v}}\right) \boldsymbol{v}.$$

That is, the projection is the scalar multiple $c\boldsymbol{v}$, where c is the ratio $\left(\frac{\boldsymbol{w} \cdot \boldsymbol{v}}{\boldsymbol{v} \cdot \boldsymbol{v}}\right)$.

For example, let $\boldsymbol{v} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$.

Then



(1) $\text{proj}_{\boldsymbol{v}}(\boldsymbol{w}) =$ ______________________________

(2) $\boldsymbol{w} - \text{proj}_{\boldsymbol{v}}(\boldsymbol{w}) =$ ______________________________

(3) The inner product of (1) and (2) is zero, as demonstrated below:

Orthogonal projection in $\boldsymbol{k}$-space. Let $\boldsymbol{v}$ and $\boldsymbol{w}$ be vectors in $\mathbf{R}^k$, and let $V = \text{Span}\{\boldsymbol{v}\}$. Then the *(orthogonal) projection* of $\boldsymbol{w}$ onto $\boldsymbol{v}$ (equivalently, the projection of $\boldsymbol{w}$ onto the subspace spanned by $\boldsymbol{v}$) is defined as follows:

$$\text{proj}_{\boldsymbol{v}}(\boldsymbol{w}) = \text{proj}_V(\boldsymbol{w}) = \left(\frac{\boldsymbol{w}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\right)\boldsymbol{v}.$$

The projection, $\widehat{\boldsymbol{w}} = \text{proj}_{\boldsymbol{v}}(\boldsymbol{w}) = \text{proj}_V(\boldsymbol{w})$, is an element of the vector space V and satisfies the following properties:

1. *Minimum Distance:* $\widehat{\boldsymbol{w}}$ is the unique vector in V *closest* to $\boldsymbol{w}$.
2. *Orthogonality:* $\widehat{\boldsymbol{w}}$ is the unique vector in V for which $\boldsymbol{w} - \widehat{\boldsymbol{w}}$ is orthogonal to V.

Thus, we can find the distance between $\boldsymbol{w}$ and V by computing $\|\boldsymbol{w} - \widehat{\boldsymbol{w}}\|$.

Problem 6. In each case, find the distance between $\boldsymbol{w}$ and $V = \text{Span}\{\boldsymbol{v}\}$.

$$(a)\ \boldsymbol{w} = \begin{bmatrix} 0 \\ 8 \\ 4 \end{bmatrix} \text{ and } V = \text{Span}\left\{\begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix}\right\} \qquad (b)\ \boldsymbol{w} = \begin{bmatrix} 1 \\ 1 \\ 7 \end{bmatrix} \text{ and } V = \text{Span}\left\{\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}\right\}$$

Orthogonal projection onto *V*. Let $V = \text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_p}\}$ be the span of an orthogonal set of vectors (a set of mutually orthogonal nonzero vectors) in $\mathbf{R}^k$ and let $\boldsymbol{w} \in \mathbf{R}^k$. Then the *(orthogonal) projection* of $\boldsymbol{w}$ onto V is defined as follows:

$$\widehat{\boldsymbol{w}} = \text{proj}_V(\boldsymbol{w}) = c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_p\boldsymbol{v_p} \quad \text{where} \quad c_i = \frac{\boldsymbol{w}\cdot\boldsymbol{v_i}}{\boldsymbol{v_i}\cdot\boldsymbol{v_i}} \quad \text{for each } i.$$

This definition generalizes the case for projection onto the span of a single vector in $\mathbf{R}^k$, and requires that the spanning set is an orthogonal set. Note also that

1. If $V_i = \text{Span}\{\boldsymbol{v_i}\}$ for each i, then $\widehat{\boldsymbol{w}}$ is the sum of projections in each coordinate direction:

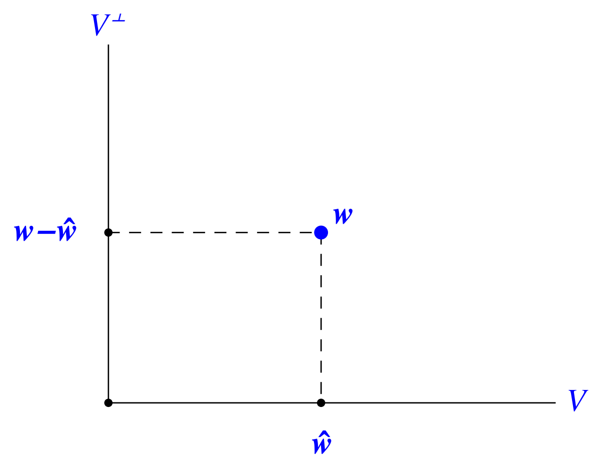
$$\widehat{\boldsymbol{w}} = \text{proj}_V(\boldsymbol{w}) = \text{proj}_{V_1}(\boldsymbol{w}) + \text{proj}_{V_2}(\boldsymbol{w}) + \cdots + \text{proj}_{V_p}(\boldsymbol{w}).$$

2. If $\boldsymbol{w} \in V$, then $\widehat{\boldsymbol{w}} = \boldsymbol{w}$.

Properties of orthogonal projections are stated in the following theorem, and illustrated to the right.

In the plot,

- the horizontal "axis" represents vector space V and the vertical "axis" represents the orthogonal complement, $V^\perp$; and
- $\boldsymbol{w}$ is decomposed into the part of $\boldsymbol{w}$ in V and the part in $V^\perp$.



Theorem (Orthogonal Projections). Let V be a subspace of $\mathbf{R}^k$, $\boldsymbol{w}$ be any vector in $\mathbf{R}^k$, and $\widehat{\boldsymbol{w}}$ be the projection of $\boldsymbol{w}$ onto V. Then

1. *Orthogonal Decomposition:* The difference $(\boldsymbol{w} - \widehat{\boldsymbol{w}})$ is a vector in $V^\perp$, and the sum

$$\boldsymbol{w} = \widehat{\boldsymbol{w}} + (\boldsymbol{w} - \widehat{\boldsymbol{w}})$$

   is the *unique* representation of $\boldsymbol{w}$ as the sum of a vector in V and a vector in $V^\perp$.

   (Thus, we have an orthogonal decomposition of $\boldsymbol{w}$ into the part of the vector in V and the part of the vector in $V^\perp$.)

2. *Best Approximation:* The vector $\widehat{\boldsymbol{w}}$ is the *closest* point in V to $\boldsymbol{w}$.

Problem 7. In each case, find $\widehat{\boldsymbol{w}}$ and $(\boldsymbol{w} - \widehat{\boldsymbol{w}})$. Note that each V has been written as the span of an orthogonal set.

$$(a)\ V = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \right\}, \ \boldsymbol{w} = \begin{bmatrix} 7 \\ 2 \\ 0 \end{bmatrix}$$

$$(b)\ V = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \\ -1 \\ 1 \end{bmatrix} \right\}, \ \boldsymbol{w} = \begin{bmatrix} 8 \\ 6 \\ 5 \\ 7 \\ -3 \end{bmatrix}$$

4.2.7 Gram-Schmidt Orthogonalization Process

Let $V \subseteq \mathbf{R}^k$ be a p-dimensional subspace and let $\{\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_p}\}$ be a basis for V. The *Gram-Schmidt orthogonalization process* allows us to construct an orthogonal basis $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_p}\}$ for V starting with $\{\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_p}\}$. The method is as follows:

1. Let $\boldsymbol{v_1} = \boldsymbol{x_1}$.
2. Let $\boldsymbol{v_2} = \boldsymbol{x_2} - \text{proj}_{V_1}(\boldsymbol{x_2})$ where $V_1 = \text{Span}\{\boldsymbol{v_1}\} = \text{Span}\{\boldsymbol{x_1}\}$.
3. Let $\boldsymbol{v_3} = \boldsymbol{x_3} - \text{proj}_{V_2}(\boldsymbol{x_3})$ where $V_2 = \text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}\} = \text{Span}\{\boldsymbol{x_1}, \boldsymbol{x_2}\}$.
4. Let $\boldsymbol{v_4} = \boldsymbol{x_4} - \text{proj}_{V_3}(\boldsymbol{x_4})$ where $V_3 = \text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\} = \text{Span}\{\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}\}$.

And, so forth. The final set, $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_p}\}$, is an orthogonal basis for V.

Problem 8. In each case, find an orthogonal basis for V.

$$(a)\ V = \text{Span}\left\{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix}\right\} \qquad (b)\ V = \text{Span}\left\{\begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 5 \\ 2 \end{bmatrix}\right\}$$

4.3 Least Squares Analysis

4.3.1 Best Approximate Solutions and the Normal Equations

Let A be an $m \times n$ coefficient matrix and assume that $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent. We propose to find *approximate* solutions to the system as follows:

(1) Find the projection of $\boldsymbol{b}$ onto $Col(A)$, $\widehat{\boldsymbol{b}}$, and

(2) Report solutions to the consistent system $A\boldsymbol{x} = \widehat{\boldsymbol{b}}$.

Observation 1: Since $\widehat{\boldsymbol{b}}$ is as close to $\boldsymbol{b}$ as possible, each approximate solution $\boldsymbol{x}$ satisfies

$$\|\boldsymbol{b} - \widehat{\boldsymbol{b}}\| = \|\boldsymbol{b} - A\boldsymbol{x}\| \text{ is as } \textit{small} \text{ as possible.}$$

The difference vector is

$$\boldsymbol{b} - A\boldsymbol{x} = \begin{bmatrix} b_1 - (a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n) \\ b_2 - (a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n) \\ \vdots \\ b_m - (a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n) \end{bmatrix}$$

and the square of the length of the difference vector is

$$\sum_{i=1}^{m} \left(b_i - (a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n)\right)^2 .$$

Each approximate solution $\boldsymbol{x}$ will minimize the above sum of squared differences. For this reason the approximate solutions are called *least squares solutions*.

Observation 2: Since the difference vector is in the orthogonal complement of the column space of A,

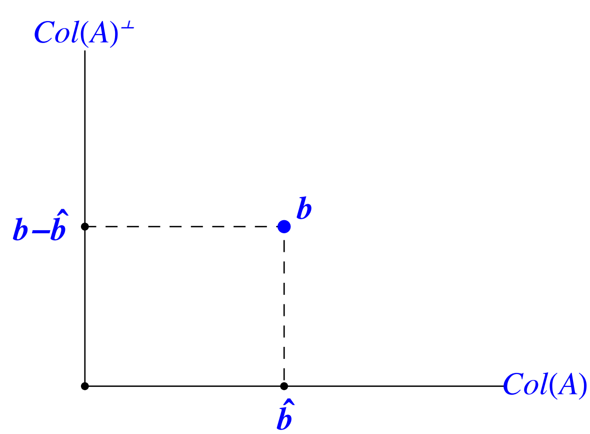
$$(\boldsymbol{b} - \widehat{\boldsymbol{b}}) = (\boldsymbol{b} - A\boldsymbol{x}) \in (Col(A))^{\perp},$$

and since the $FTLA$ implies that

$$(Col(A))^{\perp} = Null\left(A^T\right),$$

the following equality holds:

$$A^T(\boldsymbol{b} - \widehat{\boldsymbol{b}}) = A^T(\boldsymbol{b} - A\boldsymbol{x}) = \boldsymbol{O}.$$



Further,

$$A^T(\boldsymbol{b}-A\boldsymbol{x}) = \boldsymbol{O} \quad \Leftrightarrow \quad A^T\boldsymbol{b} - A^T A\boldsymbol{x} = \boldsymbol{O} \quad \Leftrightarrow \quad A^T A\boldsymbol{x} = A^T\boldsymbol{b}.$$

Thus, least squares solutions can be found by solving the consistent system on the right (called the *normal equation* of the system). By using the normal equation, we do not need to find the projection of $\boldsymbol{b}$ on the column space of A.

The following theorem gives the properties of this process:

The Least Squares Theorem. Under the conditions above,

1. $\boldsymbol{x}$ is a least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$ iff $\boldsymbol{x}$ is a solution to $A^T A\boldsymbol{x} = A^T\boldsymbol{b}$.
2. $A^T A$ is invertible iff the columns of A are linearly independent. Thus, there is a unique least squares solution iff the columns of A are linearly independent.

For example, consider the inconsistent system $A\boldsymbol{x} = \boldsymbol{b}$ where $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$. Then

$$\left[\ A^T A \mid A^T\boldsymbol{b}\ \right] = \left[\begin{array}{cc|c} 3 & 3 & 6 \\ 3 & 5 & 0 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 2 & -6 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 0 & 5 \\ 0 & 1 & -3 \end{array}\right] \Rightarrow \boldsymbol{x} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$$

is the unique least squares solution to the inconsistent system above.

Problem 1. In each case, find the least squares solution(s) to $A\boldsymbol{x} = \boldsymbol{b}$.

(a) $A = \begin{bmatrix} 1 & 4 \\ 1 & 2 \\ 2 & 3 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} -2 \\ 6 \\ 1 \end{bmatrix}$.

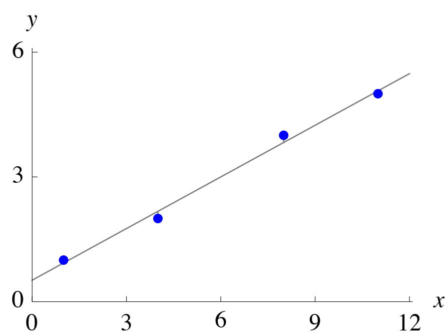
(b) $A = \begin{bmatrix} 1 & -2 \\ 2 & -4 \\ -1 & 2 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} 3 \\ -4 \\ 15 \end{bmatrix}$.

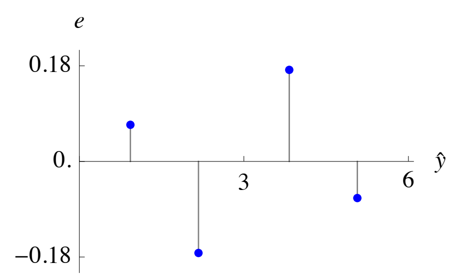
(c) $A = \begin{bmatrix} 3 & 1 & -1 \\ 1 & 2 & 0 \\ 0 & 1 & 2 \\ 1 & 1 & -1 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} -3 \\ -3 \\ 8 \\ 9 \end{bmatrix}$.

4.3.2 Application: Least Squares Analyses of Data

The methodology from the last section can be applied to finding curves of “best fit” (as minimizing a sum of squared differences).

As a simple illustration, consider the four data pairs $(1,1)$, $(4,2)$, $(8,4)$ and $(11,5)$. These points lie close to a straight line with equation $\widehat{y} = a + bx$, as illustrated in the *left* plot below.





The intercept and slope of the line can be found by the method of least squares.

Specifically, we convert the 4-by-2 system of equations to a matrix equation $A\boldsymbol{x} = \boldsymbol{b}$, and find the least squares solution(s) by solving the normal equation $A^T A\boldsymbol{x} = A^T\boldsymbol{b}$.

$$\begin{array}{rcl} a + b\,1 & = & 1 \\ a + b\,4 & = & 2 \\ a + b\,8 & = & 4 \\ a + b11 & = & 5 \end{array} \Rightarrow \begin{bmatrix} 1 & 1 \\ 1 & 4 \\ 1 & 8 \\ 1 & 11 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \\ 5 \end{bmatrix} \Rightarrow \begin{bmatrix} 4 & 24 \\ 24 & 202 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 12 \\ 96 \end{bmatrix} \Rightarrow \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 15/29 \\ 12/29 \end{bmatrix}$$

Thus, the least squares regression line is

$$\widehat{y} = \left(\tfrac{15}{29}\right) + \left(\tfrac{12}{29}\right) x \approx 0.52 + 0.41x,$$

as shown on the *left* above.

Let $\widehat{y_i} = \left(\frac{15}{29}\right) + \left(\frac{12}{29}\right) x_i$ and $e_i = y_i - \widehat{y_i}$ for $i = 1, 2, 3, 4$:

$\widehat{y_i}$	0.931	2.172	3.828	5.069
e_i	0.069	−0.172	0.172	−0.069

A plot of $(\widehat{y_i}, e_i)$ pairs is shown on the *right* above.

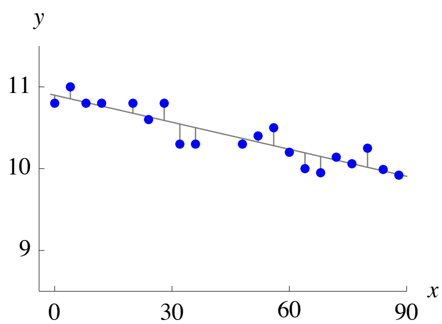
The following pages contain several applications of this general strategy to real data.

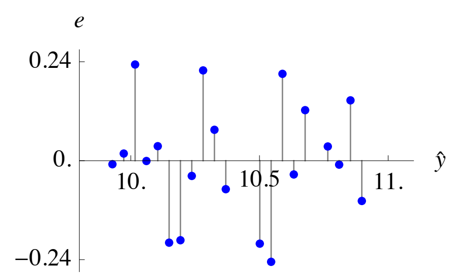
Example: Olympic winning times (***Source:*** **Hand et al, 1994).** Consider the following 20 data pairs, where x_i is the time in years since 1900 and y_i is the Olympic winning time in seconds for men in the final round of the 100 meter event.

i	1	2	3	4	5	6	7	8	9	10
x_i	0	4	8	12	20	24	28	32	36	48
y_i	10.8	11.0	10.8	10.8	10.8	10.6	10.8	10.3	10.3	10.3
i	11	12	13	14	15	16	17	18	19	20
x_i	52	56	60	64	68	72	76	80	84	88
y_i	10.4	10.5	10.2	10.0	9.95	10.14	10.06	10.25	9.99	9.92

The data cover all Olympic events held between 1900 and 1988. (Note that Olympic games were not held in 1916, 1940, and 1944.)

The twenty data pairs lie approximately on a straight line with equation $\widehat{y} = a + bx$, whose intercept and slope can be estimated by the method of least squares.





$$\begin{array}{rcr} a + b\,0 & = & 10.80 \\ a + b\,4 & = & 11.00 \\ \cdots\cdots & \cdots & \cdots \\ a + b 88 & = & 9.92 \end{array} \Rightarrow \begin{bmatrix} 1 & 0 \\ 1 & 4 \\ \cdots & \cdots \\ 1 & 88 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 10.80 \\ 11.00 \\ \cdots \\ 9.92 \end{bmatrix} \Rightarrow \begin{bmatrix} 20 & 912 \\ 912 & 56928 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 207.91 \\ 9311.76 \end{bmatrix}$$

which implies that $\begin{bmatrix} a \\ b \end{bmatrix} \approx \begin{bmatrix} 10.898 \\ -0.011 \end{bmatrix}$. Thus, the least squares regression line is

$$\widehat{y} = 10.898 - 0.011x,$$

as illustrated on the *left* above. Note that the origin of the plot is not $(0, 0)$.

The results suggest that the winning times have decreased at the rate of about 0.011 seconds per year during the 88 years of the study.

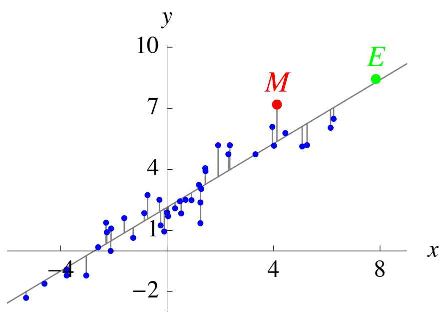
Let $\widehat{y_i} = 10.898 - 0.011x_i$ and $e_i = y_i - \widehat{y_i}$ for $i = 1, 2, \ldots, 20$.

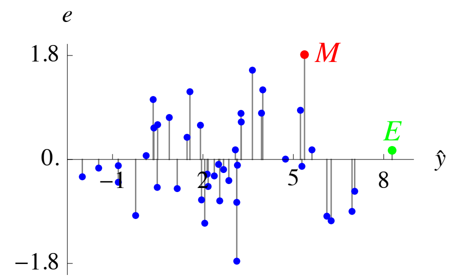
A plot of $(\widehat{y_i}, e_i)$ pairs is shown on the *right* above.

Example: Brain-body study (***Source:*** **Allison & Cicchetti, 1976).** As part of a study on sleep in mammals, researchers collected information on the average body weight (in kilograms) and average brain weight (in grams) for 43 different species. Let

$$x_i = \ln(\text{Average Body Weight}_i) \quad \text{and} \quad y_i = \ln(\text{Average Brain Weight}_i)$$

for $i = 1, 2, \ldots, 43$. The (x_i, y_i) pairs lie approximately on a line with equation $\widehat{y} = a + bx$, whose intercept and slope can be estimated by the method of least squares.





Starting with the normal equation $A^T A\boldsymbol{x} = A^T \boldsymbol{b}$:

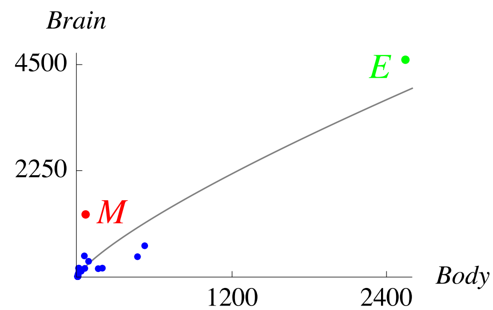
$$\begin{bmatrix} 43 & 30.8655 \\ 30.8655 & 416.725 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 116.26 \\ 392.109 \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} a \\ b \end{bmatrix} \approx \begin{bmatrix} 2.142 \\ 0.782 \end{bmatrix}.$$

Thus, the least squares regression line is $\widehat{y} = 2.142 + 0.782x$.

As with the earlier examples, the *left* plot above is a plot of (x_i, y_i)-pairs superimposed on the least squares regression line and the *right* plot is a plot of $(\widehat{y}_i, e_i)$ pairs, for $i = 1, 2, \ldots, 43$.

It is instructive to examine the estimated relationship between average brain and body weights on their original scales. The graph of this relationship is shown to the right.

The formula for this curve is (please complete)



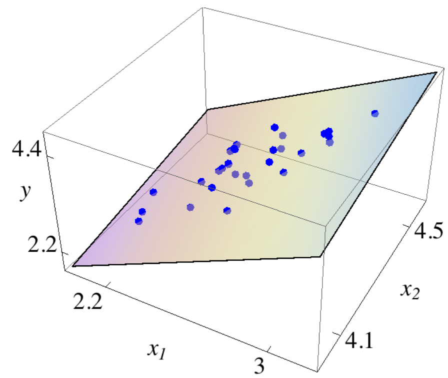
____.

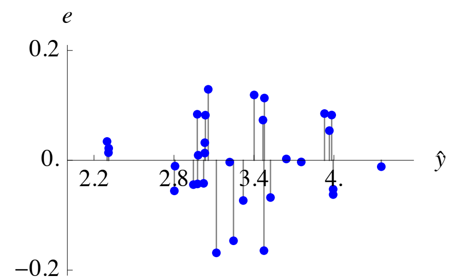
Note that *Man*'s brain weight is much larger than expected given the modest body weight. The Asian *Elephant* has an enormous body weight and a correspondingly large brain weight.

Example: Timber yield study (***Source:*** **Hand et al, 1994).** As part of a study designed to estimate the volume of a tree (and therefore its yield) given its diameter and height, data were collected on the volume (in cubic feet), diameter at 54 inches above the ground (in inches), and height (in feet) of 31 black cherry trees in the Allegheny National Forest. Let

$$x_{1,i} = \ln(\text{Diameter}_i), \quad x_{2,i} = \ln(\text{Height}_i) \;\text{ and }\; y_i = \ln(\text{Volume}_i) \text{ for } i = 1, \ldots, 31.$$

The $(x_{1,i}, x_{2,i}, y_i)$ triples lie approximately on a plane with equation $\widehat{y} = a + bx_1 + cx_2$, whose coefficients can be estimated using the method of least squares.





Starting with the normal equation $A^T A\boldsymbol{x} = A^T\boldsymbol{b}$:

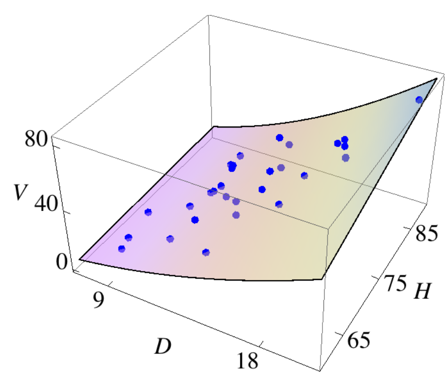
$$\begin{bmatrix} 31 & 79.278 & 134.144 \\ 79.278 & 204.376 & 343.37 \\ 134.144 & 343.37 & 580.694 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 101.455 \\ 263.056 \\ 439.896 \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} a \\ b \\ c \end{bmatrix} \approx \begin{bmatrix} -6.632 \\ 1.983 \\ 1.117 \end{bmatrix}.$$

Thus, the least squares regression equation is $\widehat{y} = -6.632 + 1.983x_1 + 1.117x_2$.

The regression equation is plotted on the *left* above, along with the $(x_{i,1}, x_{2,i}, y_i)$ triples. Triples lying under the surface appear slightly lighter in color. The *right* plot is a plot of $(\widehat{y}_i, e_i)$ pairs, for $i = 1, 2, \ldots, 31$.

It is instructive to examine the estimated relationship among diameter, height and volume in their original scales. The graph of this relationship is shown to the right.

The formula for this curve is (please complete)



____.

Any comments?

Example: Body fat study (***Source:*** **Johnson, 1996**). As part of a study to determine if the percentage of body fat can be predicted accurately using only a scale and measuring tape, data were collected on 100 men. Let $x_{1,i}$, $x_{2,i}$, $x_{3,i}$ and $x_{4,i}$ be the abdomen, wrist, hip and neck circumferences (in centimeters) of the i^{th} individual, and let y_i be the man's percent body fat measured using an accurate underwater technique. Here is some summary information:

	Average Value	Minimum Value	Maximum Value
Abdomen (x_1)	93.365	74.6	148.1
Wrist (x_2)	18.246	16.1	21.4
Hip (x_3)	100.163	85.3	147.7
Neck (x_4)	38.174	31.1	51.2
Body Fat (y)	19.444	0.7	47.5

Consider fitting a linear function of the form $\widehat{y} = a + bx_1 + cx_2 + dx_3 + ex_4$ using the method of least squares to estimate the coefficients (a through e).

Starting with the normal equation $A^T A\boldsymbol{x} = A^T\boldsymbol{b}$:

$$\begin{bmatrix} 100.0 & 9336.5 & 1824.6 & 10016.3 & 3817.4 \\ 9336.5 & 885594.7 & 171109.9 & 943738.8 & 359012.2 \\ 1824.6 & 171109.9 & 33388.4 & 183261.6 & 69843.1 \\ 10016.3 & 943738.8 & 183261.6 & 1009635.7 & 384080.0 \\ 3817.4 & 359012.2 & 69843.1 & 384080.0 & 146454.6 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix} = \begin{bmatrix} 1944.4 \\ 190340.0 \\ 35815.9 \\ 199664.6 \\ 75630.6 \end{bmatrix} \Rightarrow \begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix} \approx \begin{bmatrix} 9.88 \\ 1.04 \\ -1.89 \\ -0.36 \\ -0.45 \end{bmatrix}$$

Thus, the least squares regression formula is $\widehat{y} = 9.88 + 1.04x_1 - 1.89x_2 - 0.36x_3 - 0.45x_4$.

The $(\widehat{y}_i, e_i)$ pairs are shown on the *left* below. Individuals with

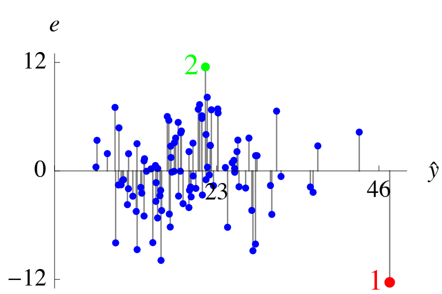
1. the largest $\widehat{y}_i$ and smallest e_i, and measurements

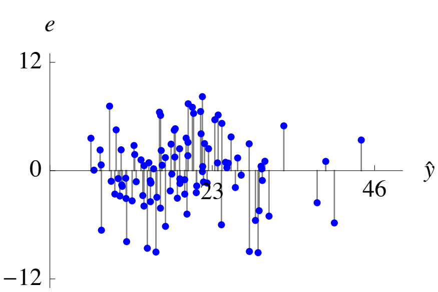
   (Abdomen, Wrist, Hip, Neck, Body Fat) = (148.1, 21.4, 147.7, 51.2, 35.2), and

2. the largest e_i, and measurements

   (Abdomen, Wrist, Hip, Neck, Body Fat) = (93.9, 17.3, 100.1, 39.1, 32.9),

have been highlighted in the plot. Any comments?





If these two individuals are removed and a new least squares solution is computed, the pattern of errors does not change dramatically, as shown in the *right* plot above.

4.3.3 Footnote: Eigenvalues, Eigenvectors and Least Squares Analysis

Let $M = A^T A$ be the matrix used in the normal equation for finding least squares solutions to inconsistent systems. M is a symmetric matrix, satisfying the following properties:

1. M is a diagonalizable matrix,
2. the eigenvalues of M are nonnegative real numbers,
3. M has an eigenvector basis that is an orthogonal set, and
4. M is invertible if and only if all eigenvalues are positive.

If M is invertible, then the least squares solution is unique. Further, as long as the smallest eigenvalue is not "too close to zero," then the computer will have no trouble finding the unique solution accurately.

Body fat study example, continued. Consider again the body fat study from the last section. $M = A^T A$ can be written as $M = PDP^{-1}$, where

$$D \approx \begin{bmatrix} 2072726.621 & 0 & 0 & 0 & 0 \\ 0 & 2059.74 & 0 & 0 & 0 \\ 0 & 0 & 330.145 & 0 & 0 \\ 0 & 0 & 0 & 56.593 & 0 \\ 0 & 0 & 0 & 0 & 0.198 \end{bmatrix} \text{ and}$$

$$P \approx \begin{bmatrix} -0.007 & -0.017 & 0.015 & -0.031 & 0.999 \\ -0.653 & 0.753 & 0.063 & -0.052 & 0.005 \\ -0.127 & -0.201 & 0.332 & -0.912 & -0.037 \\ -0.698 & -0.564 & -0.437 & 0.063 & -0.006 \\ -0.265 & -0.273 & 0.834 & 0.400 & -0.007 \end{bmatrix}.$$

The eigenvalues of M are written in decreasing order along the diagonal of D; all eigenvalues are "comfortably" greater than zero , implying that the computer had no trouble finding accurate least squares estimates of the coefficients of the prediction formula.

Finally, to improve the accuracy of least squares estimates in situations where eigenvalues may be close to zero (and M may be "close to singular"), practitioners use a *singular value decomposition* of M before trying to find the estimates.