Eigenvector

>> of an $n \times n$ matrix A is a non-zero vector x such that $Ax = \lambda x$ for some scalar λ.

direction of image of A does not change; x & Ax points to same direction

Eigenvalue

>> A scalar λ if there is a non-trivial solution x of $Ax = \lambda x$; such an x is called an 'eigenvector corresponding to λ'

* eigen vector cannot be 0, → to avoid $A0 = \lambda 0$, which holds for every A and λ

eigen value can be 0,

they are only for SQUARE matrixs.

EXAMPLE 2 Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$. Are $\mathbf{u}$ and $\mathbf{v}$ eigenvectors of A?

↗ v is not an eigenvector of A

SOLUTION

$$A\mathbf{u} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} 6 \\ -5 \end{bmatrix} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = -4\begin{bmatrix} 6 \\ -5 \end{bmatrix} = -4\mathbf{u}$$

↙ multiple of vector ∴ eigen vector

$$A\mathbf{v} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 11 \end{bmatrix} \neq \lambda\begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

← not multiples of each other ∴ not eigen vector

Thus $\mathbf{u}$ is an eigenvector corresponding to an eigenvalue (-4), but $\mathbf{v}$ is not an eigenvector of A, because $A\mathbf{v}$ is not a multiple of $\mathbf{v}$. ■

↑ scalar quantity is known as the eigen value

Find eigen vectors if eigenvalue unknown

EXAMPLE 3 Show that 7 is an eigenvalue of matrix A in Example 2, and find the corresponding eigenvectors.

SOLUTION The scalar 7 is an eigenvalue of A if and only if the equation

$$A\mathbf{x} = 7\mathbf{x} \tag{1}$$

has a nontrivial solution. But (1) is equivalent to $A\mathbf{x} - 7\mathbf{x} = \mathbf{0}$, or

$$(A - 7I)\mathbf{x} = \mathbf{0} \tag{2}$$

To solve this homogeneous equation, form the matrix

$$A - 7I = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} - \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} -6 & 6 \\ 5 & -5 \end{bmatrix}$$

linearly dependent set ↓ hence got non-trivial soln

The columns of $A - 7I$ are obviously linearly dependent, so (2) has nontrivial solutions. Thus 7 *is* an eigenvalue of A. To find the corresponding eigenvectors, use row operations:

$$\begin{bmatrix} -6 & 6 & 0 \\ 5 & -5 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The general solution has the form $x_2\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ ← eigenvector. Each vector of this form with $x_2 \neq 0$ is an eigenvector corresponding to $\lambda = 7$. ■

↖ eigen value

EigenSpace

The equivalence of equations (1) and (2) obviously holds for any λ in place of $\lambda = 7$. Thus λ is an eigenvalue of an $n \times n$ matrix A if and only if the equation

$$(A - \lambda I)\mathbf{x} = \mathbf{0} \tag{3}$$

→ eigen vector → tied to λ

has a nontrivial solution. The set of *all* solutions of (3) is just the null space of the matrix $A - \lambda I$. So this set is a *subspace* of $\mathbb{R}^n$ and is called the **eigenspace** of A corresponding to λ. The eigenspace consists of the zero vector and all the eigenvectors corresponding to λ.

>> eigen value could be repeated and there could be >1 independent eigen vector

EigenSpace: Example two-dimensional subspace of R^3

Example:

In this example, there is **repeated eigenvalue.**

When this occur, it is **possible to have > 1 eigenvector for that single eigenvalue.**

I.e, the eigenspace's dimension is greater than 1.

EXAMPLE 4 Let $A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$. An eigenvalue of A is 2. Find a basis for the corresponding eigenspace.

SOLUTION Form

$$A - 2I = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix} - \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 6 \end{bmatrix}$$

and row reduce the augmented matrix for $(A - 2I)\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} 2 & -1 & 6 & 0 \\ 2 & -1 & 6 & 0 \\ 2 & -1 & 6 & 0 \end{bmatrix} \sim \begin{bmatrix} 2 & -1 & 6 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

x_2 & x_3 free variables

At this point, it is clear that 2 is indeed an eigenvalue of A because the equation $(A - 2I)\mathbf{x} = \mathbf{0}$ has free variables. The general solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2\begin{bmatrix} 1/2 \\ 1 \\ 0 \end{bmatrix} + x_3\begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}, \quad x_2 \text{ and } x_3 \text{ free}$$

The eigenspace, shown in Fig. 3, is a two-dimensional subspace of $\mathbb{R}^3$. A basis is

$$\left\{\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}\right\}$$ ■

Steps

To find eigenvalues & vectors:

1. find eigen value of matrix
2. Gaussian Elimination (REF) to find solution of homogenous equation $(A - \lambda I)x = 0$ for each eigenvalue λ ↳ null space of x

Finding eigenvalues

$Ax = \lambda x \Leftrightarrow Ax = AIx \Leftrightarrow (\lambda I - A)x = 0$

x cannot be a zero vector

$(\lambda I - A)$ must be a singular matrix

→ $\det(A - \lambda I) = 0$

→ cols & rows of A are linearly independent

>> linear dependence obvious when:

2 rows / cols are the same OR

row / col is all 0

Characteristic Eqn & Polynomial

EXAMPLE 1 Find the eigenvalues of $A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$.

SOLUTION We must find all scalars λ such that the matrix equation

$$(A - \lambda I)\mathbf{x} = \mathbf{0}$$

has a nontrivial solution. By the Invertible Matrix Theorem in Section 2.3, this problem is equivalent to finding all λ such that the matrix $A - \lambda I$ is *not* invertible, where

$$A - \lambda I = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix}$$

By Theorem 4 in Section 2.2, this matrix fails to be invertible precisely when its determinant is zero. So the eigenvalues of A are the solutions of the equation

$$\det(A - \lambda I) = \det\begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix} = 0$$

Recall that

$$\det\begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

So

Characteristic Eqn ↘

$$\boxed{\det(A - \lambda I)} = (2-\lambda)(-6-\lambda) - (3)(3)$$
$$= -12 + 6\lambda - 2\lambda + \lambda^2 - 9$$
$$\boxed{= \lambda^2 + 4\lambda - 21} \rightarrow \text{Characteristic Polynomial}$$
$$= (\lambda - 3)(\lambda + 7)$$

If $\det(A - \lambda I) = 0$, then $\lambda = 3$ or $\lambda = -7$. So the eigenvalues of A are 3 and -7. ■

Eigenvalue == 0

if a matrix A has an eigenvalue of 0,

$Ax = 0x$ has a non-trivial solution.

NOT POSSIBLE! Unless.

$Ax = 0$ has a non-trivial solution,

but A is NOT invertible.

∴ 0 is an eigenvalue of $A \Leftrightarrow A$ is not invertible.

Theorem

square matrix ↙

Given $Ax = \lambda x$:

>> λ is an eigenvalue of A,

>> $N(A - \lambda I) \neq \{0\}$

>> matrix $A - \lambda$ is singular

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

EigenValues of Δ Matrices Theorem

eigenvalues of Δ matrix are the entries on its main diagonal

Fact: determinant of diagonal matrix is the product of diagonal entries.

Invertible Matrix Theorem (continued)

A: $n \times n$ matrix, invertible if and only if:

>> 0 is NOT an eigenvalue of A

>> $\det$ of $A \neq 0$

Example: eigenvalue and algebraic multiplicity

EXAMPLE 3
Find the characteristic equation of

$$A = \begin{bmatrix} 5 & -2 & 6 & -1 \\ 0 & 3 & -8 & 0 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

SOLUTION Form $A - \lambda I$, and use Theorem 3(d):

$$\det(A - \lambda I) = \det\begin{bmatrix} 5-\lambda & -2 & 6 & -1 \\ 0 & 3-\lambda & -8 & 0 \\ 0 & 0 & 5-\lambda & 4 \\ 0 & 0 & 0 & 1-\lambda \end{bmatrix}$$
$$= (5-\lambda)(3-\lambda)(5-\lambda)(1-\lambda)$$

The characteristic equation is

$$(5-\lambda)^2(3-\lambda)(1-\lambda) = 0$$

or

$$(\lambda-5)^2(\lambda-3)(\lambda-1) = 0$$

Expanding the product, we can also write

$$\lambda^4 - 14\lambda^3 + 68\lambda^2 - 130\lambda + 75 = 0$$

The eigenvalue 5 in Example 3 is said to have *multiplicity* 2 because $(\lambda - 5)$ occurs two times as a factor of the characteristic polynomial. In general, the (**algebraic**) **multiplicity** of an eigenvalue λ is its multiplicity as a root of the characteristic equation.

Lay,4th,pg276

EigenValues and Algebraic multiplicity

no constant; easy to factorize ↓

EXAMPLE 4 The characteristic polynomial of a 6×6 matrix is $\lambda^6 - 4\lambda^5 - 12\lambda^4$. Find the eigenvalues and their multiplicities.

SOLUTION Factor the polynomial

$$\lambda^6 - 4\lambda^5 - 12\lambda^4 = \lambda^4(\lambda^2 - 4\lambda - 12) = \lambda^4(\lambda - 6)(\lambda + 2)$$

The eigenvalues are 0 (multiplicity 4), 6 (multiplicity 1), and -2 (multiplicity 1). ■

The **set of solutions** $(\lambda_1, \lambda_2, \ldots, \lambda_{N_\lambda})$, that is, the **eigenvalues**, is called the **spectrum** of A. The charaacteristic polynomial can be factored as follows:

$$p(\lambda) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_{N_\lambda})^{n_{N_\lambda}} = 0.$$

The integer n_i is termed the **algebraic multiplicity** of eigenvalue λ_i. It is the number of times an eigenvalue appears as a root of the characteristic polynomial.

The algebraic multiplicities sum to N (the number of rows in A):

$$\sum_{i=1}^{N_\lambda} n_i = N.$$

For each eigenvalue λ_i, there is a corresponding EigenSpace $E(\lambda_i)$

Similarity & Diagonalization

$n \times n$ matrixes A & P, P is invertible:

$$P^{-1}AP = B$$
$$AP = PB$$
$$A = PBP^{-1}$$

→ A & B are similar matrices

→ transformation from A to $B = P^{-1}AP$ is called similarity transformation.

→ if B is a diagonal matrix, A is diagonalizable!

Table 1 Similarity Invariants

Property	Description
Determinant	A and $P^{-1}AP$ have the same determinant.
Invertibility	A is invertible if and only if $P^{-1}AP$ is invertible.
Rank	A and $P^{-1}AP$ have the same rank.
Nullity	A and $P^{-1}AP$ have the same nullity.
Trace	A and $P^{-1}AP$ have the same trace.
Characteristic polynomial	A and $P^{-1}AP$ have the same characteristic polynomial.
Eigenvalues	A and $P^{-1}AP$ have the same eigenvalues.
Eigenspace dimension	If λ is an eigenvalue of A (and hence of $P^{-1}AP$) then the eigenspace of A corresponding to λ and the eigenspace of $P^{-1}AP$ corresponding to λ have the same dimension.

Properties of diagonal matrix

>> eigenvalues of diagonal matrixes are its diagonal element

>> determinant = product of diagonal entries

>> Rank = no. of non-zero entries in diagonal

4) Multiplication: given A and diagonal matrix D (AD and DA):

- when we pre-multiply A by a diagonal matrix D, the rows of A are multiplied by the diagonal elements of D;
- when we post-multiply A by D, the columns of A are multiplied by the diagonal elements of D.

>> diagonal matrix's inverse is reciprocal of diagonal elements

>> Product is easy to compute

Question: When is A diagonalizable?

Ans: (Refer to Theorem 5)

A is diagonalizable ⟺ there are enough eigenvectors to form a basis of $\mathbb{R}^n$

→ eigenvector basis of $\mathbb{R}^n$

when A has n independent eigenvectors.

Theorem 5

The Diagonalization Theorem

An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

In fact, $A = PDP^{-1}$, with D a diagonal matrix, if and only if the columns of P are n linearly independent eigenvectors of A. In this case, the diagonal entries of D are eigenvalues of A that correspond, respectively, to the eigenvectors in P.

2. Let $A = \begin{bmatrix} -3 & 12 \\ -2 & 7 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, and $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Suppose you are told that $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors of A. Use this information to diagonalize A.

Sol:

2. Compute $A\mathbf{v}_1 = \begin{bmatrix} -3 & 12 \\ -2 & 7 \end{bmatrix}\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 1 \cdot \mathbf{v}_1$, and

$$A\mathbf{v}_2 = \begin{bmatrix} -3 & 12 \\ -2 & 7 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix} = 3 \cdot \mathbf{v}_2$$

So, $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors for the eigenvalues 1 and 3, respectively. Thus

$$A = PDP^{-1}, \quad \text{where} \quad P = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$$

Steps

Given a matrix A size nxn, to diagonalize it to D, perform the following:

1) First find the eigenvalues of A (solve $\det(A - \lambda I) = 0$).
2) For each eigenvalue λ_i, find the eigenvectors of corresponding λ_i
3) **If** there are n independent eigenvectors v_i, **then** the matrix A can be represented as:

$$AP = PD$$
$$A = PDP^{-1}$$
$$P^{-1}AP = D$$

Where D = diagonal matrix with eigenvalues λ_i
And P is the matrix with columns that **are corresponding eigenvectors v_i.**

Theorem 6

An $n \times n$ matrix with n distinct eigenvalues is diagonizable.

Theorem 2

if $v_1, \dots, v_r$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, \dots, \lambda_r$ of an $n \times n$ matrix A, then the set $\{v_1 \dots v_r\}$ is linearly independent

key: distinct eigenvalues → distinct eigenvectors that are LI.

Example: Distinct EigenValues -> Diagonalizable

EXAMPLE 5 Determine if the following matrix is diagonalizable.

$$A = \begin{bmatrix} 5 & -8 & 1 \\ 0 & 0 & 7 \\ 0 & 0 & -2 \end{bmatrix}$$

SOLUTION This is easy! Since the matrix is triangular, its eigenvalues are obviously 5, 0, and -2. Since A is a 3×3 matrix with three distinct eigenvalues, A is diagonalizable. ■

See Slide 8.1.2 (pg 8) To see slide "EigenValues of Triangular Matrixes"

THEOREM 1 The eigenvalues of a triangular matrix are the entries on its main diagonal.

Distinct eigenvalue is a sufficient BUT not necessary condition to have linearly independent eigenvectors.

Example 3 (later) shows that eigenvalues are repeated, but it is still diagonalizable.
And Example 4 shows counter-example.

When is a matrix with repeated roots diagonalizable? Introducing algebraic and geometric multiplicity.

Algebraic multiplicity = multiplicity of eigenvalues λ_k

Geometric multiplicity = Dimension of eigenspace corresponding to eigenvalue λ_k

Ref: https://people.math.carleton.ca/~kcheung/math/notes/MATH1107/wk10/10_algebraic_and_geometric_multiplicities.html

Introducing terminology: Algebraic and Geometric Multiplicity

Eigenspaces

Let λ be an eigenvalue of A. Recall that the eigenvectors of A for λ are the nonzero vectors in the nullspace of $A - \lambda I$. We call the nullspace $A - \lambda I$ the **eigenspace** of A for λ denoted by $\mathcal{E}_A(\lambda)$. In other words, $\mathcal{E}_A(\lambda)$ consists of all the eigenvectors of A for λ and the zero vector.

Example

Let $A = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}$. Note that -1 is an eigenvalue of A. Then $A - (-1)I_2 = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$. The nullspace of this matrix is spanned by the single vector $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Hence, $\mathcal{E}_A(-1)$ is the span of $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

Sanity check:

```
% example in sli
A = [1 2; 1 0]
[P,D] = eig(A)
```

```
A =

     1     2
     1     0


P =

    0.8944   -0.7071
    0.4472    0.7071


D =

     2     0
     0    -1
```

Ref:https://people.math.carleton.ca/~kcheung/math/notes/MATH1107/wk10/10_algebraic_and_geometric_multiplicities.html

Geometric Multiplicity (is the dimension of Eigen Space) Algebraic Multiplicity (is the number of repeated roots)

Algebraic multiplicity vs geometric multiplicity

The **geometric multiplicity** of an eigenvalue λ of A is the dimension of $\mathcal{E}_A(\lambda)$. In the example above, the geometric multiplicity of -1 is 1 as the eigenspace is spanned by one nonzero vector.

In general, determining the geometric multiplicity of an eigenvalue requires no new technique because one is simply looking for the dimension of the nullspace of $A - \lambda I$.

The **algebraic multiplicity** of an eigenvalue λ of A is the number of times λ appears as a root of p_A. For the example above, one can check that -1 appears only once as a root. Let us now look at an example in which an eigenvalue has multiplicity higher than 1.

In mathematics, the **dimension** of a vector space *V* is the cardinality (i.e. the number of vectors) of a basis of *V* over its base field.[1]

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

```
% Example:
% repeated roots lambda=1
% BUT eigenspace ==1
A = [1 2; 0 1]
[P,D] = eig(A)
```

```
A =

     1     2
     0     1


P =

    1.0000   -1.0000
         0    0.0000


D =

     1     0
     0     1
```

Ref:https://people.math.carleton.ca/~kcheung/math/notes/MATH1107/wk10/10_algebraic_and_geometric_multiplicities.html

When is a matrix with repeated roots diagonalizable? Introducing algebraic and geometric multiplicity.

In general, **the algebraic multiplicity and geometric multiplicity of an eigenvalue can differ. However, the geometric multiplicity can never exceed the algebraic multiplicity**.

It is a fact that summing up the algebraic multiplicities of all the eigenvalues of an $n \times n$ matrix A gives exactly n. **If for every eigenvalue of A, the geometric multiplicity equals the algebraic multiplicity, then A is said to be *diagonalizable*.**

See also (Theorem 7) in Lay, 4thEd, pg 285, Ch 5.3

Ref: https://people.math.carleton.ca/~kcheung/math/notes/MATH1107/wk10/10_algebraic_and_geometric_multiplicities.html

proof not tested!

Example 3: Diagonalizable A with repeated eigenValue

EXAMPLE 3 Diagonalize the following matrix, if possible.

$$A = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}$$

That is, find an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$.

SOLUTION There are four steps to implement the description in Theorem 5.

Step 1. Find the eigenvalues of A. As mentioned in Section 5.2, the mechanics of this step are appropriate for a computer when the matrix is larger than 2×2. To avoid unnecessary distractions, the text will usually supply information needed for this step. In the present case, the characteristic equation turns out to involve a cubic polynomial that can be factored:

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

The eigenvalues are $\lambda = 1$ and $\lambda = -2$.

Note: In this example, the eigenvalues are NOT distinct ($\lambda = -2$), i.e repeated, But this matrix is diagonalizable.

Algebraic multiplicity of eigenvalue = -2 is 2
AND
Geometric multiplicity of eigenvalue = -2 is ALSO 2.

HENCE there is a complete set of linearly independent eigenvectors for A, allowing A to be diagonalizable.

Lay4th, pg 283, Ch 5.2

Example 3: Diagonalizable A with repeated eigenValue

Step 2. Find three linearly independent eigenvectors of A. Three vectors are needed because A is a 3×3 matrix. This is the critical step. If it fails, then Theorem 5 says that A cannot be diagonalized. The method in Section 5.1 produces a basis for each eigenspace:

$$\text{Basis for } \lambda = 1: \quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$\text{Basis for } \lambda = -2: \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

You can check that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly independent set.

Step 3. Construct P from the vectors in step 2. The order of the vectors is unimportant. Using the order chosen in step 2, form

$$P = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3] = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

eigenvector

Sanity check:

```
% Example: slide 5.1.3, pg 17
A = [1 3 3; -3 -5 -3; 3 3 1]
[P,D] = eig(A)
```

```
A =

     1     3     3
    -3    -5    -3
     3     3     1

P =

   -0.5774   -0.7876    0.4206
    0.5774    0.2074   -0.8164
   -0.5774    0.5802    0.3957

D =

    1.0000         0         0
         0   -2.0000         0
         0         0   -2.0000
```

```
P23 = P(:,2:3);
v2 = [-1 1 0]'
c2 = pinv(P23)*v2
v2_est = P23*c2
```

```
c2 =

    0.7121
   -1.0440

v2_est =

   -1.0000
    1.0000
   -0.0000
```

Note that matlab produces col 2,3 of P that does not resemble v_2 *and* v_3.
But you can check that v_2 *and* v_3 can be formed by appropriate linear combinations of col2,3, of P!

Example 3: Diagonalizable A with repeated eigenValue

Step 4. Construct D from the corresponding eigenvalues. In this step, it is essential that the order of the eigenvalues matches the order chosen for the columns of P. Use the eigenvalue $\lambda = -2$ twice, once for each of the eigenvectors corresponding to $\lambda = -2$:

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

It is a good idea to check that P and D really work. To avoid computing P^{-1}, simply verify that $AP = PD$. This is equivalent to $A = PDP^{-1}$ when P is invertible. (However, be sure that P is invertible!) Compute

$$AP = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 \\ -1 & -2 & 0 \\ 1 & 0 & -2 \end{bmatrix}$$

$$PD = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 \\ -1 & -2 & 0 \\ 1 & 0 & -2 \end{bmatrix}$$

same

Example 4: NOT Diagonalizable A with repeated eigenValue

EXAMPLE 4 Diagonalize the following matrix, if possible.

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

SOLUTION The characteristic equation of A turns out to be exactly the same as that in Example 3:

$$0 = \det(A - \lambda I) = -\lambda^3 - 3\lambda^2 + 4 = -(\lambda - 1)(\lambda + 2)^2$$

The eigenvalues are $\lambda = 1$ and $\lambda = -2$. However, it is easy to verify that each eigenspace is only one-dimensional:

$$\text{Basis for } \lambda = 1: \quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$\text{Basis for } \lambda = -2: \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

There are no other eigenvalues, and every eigenvector of A is a multiple of either $\mathbf{v}_1$ or $\mathbf{v}_2$. Hence it is impossible to construct a basis of $\mathbb{R}^3$ using eigenvectors of A. By Theorem 5, A is *not* diagonalizable.

It is not possible to diagonalize as A does not have a full set of independent eigenvectors.

Note: In this example, the eigenvalues are NOT distinct ($\lambda = -2$), i.e repeated, But this matrix is NOT diagonalizable.

Algebraic multiplicity of eigenvalue = -2 is 2
BUT
Geometric multiplicity of eigenvalue = -2 is ONLY 1.

Hence incomplete basis of eigenvectors for A =>
A is NOT diagonalizable.

Power of A: Example 1

In many cases, the eigenvalue–eigenvector information contained within a matrix A can be displayed in a useful factorization of the form $A = PDP^{-1}$ where D is a diagonal matrix. In this section, the factorization enables us to compute A^k quickly for large values of k, a fundamental idea in several applications of linear algebra. Later, in Sections 5.6 and 5.7, the factorization will be used to analyze (and *decouple*) dynamical systems.

The following example illustrates that powers of a diagonal matrix are easy to compute.

EXAMPLE 1 If $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$, then $D^2 = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix}$

and

$$D^3 = DD^2 = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix} = \begin{bmatrix} 5^3 & 0 \\ 0 & 3^3 \end{bmatrix}$$

In general,

$$D^k = \begin{bmatrix} 5^k & 0 \\ 0 & 3^k \end{bmatrix} \quad \text{for } k \geq 1$$

Lay, pg 281

Example 2: Finding A^k from $A = PDP^{-1}$

EXAMPLE 2 Let $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. Find a formula for A^k, given that $A = PDP^{-1}$, where

$$P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$$

SOLUTION The standard formula for the inverse of a 2×2 matrix yields

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

Then, by associativity of matrix multiplication,

$$A^2 = (PDP^{-1})(PDP^{-1}) = PD\underbrace{(P^{-1}P)}_{I}DP^{-1} = PDDP^{-1}$$

$$= PD^2P^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

Note: P is formed by the 2 eigenvectors of A! 5 & 3 in D are the eigenvalues of A!

Example 2: Finding A^k from $A = PDP^{-1}$

Again,

$$A^3 = (PDP^{-1})A^2 = (PD\underbrace{P^{-1})PD^2}_{I}P^{-1} = PDD^2P^{-1} = PD^3P^{-1}$$

In general, for $k \geq 1$,

$$A^k = PD^kP^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} 5^k & 0 \\ 0 & 3^k \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \cdot 5^k - 3^k & 5^k - 3^k \\ 2 \cdot 3^k - 2 \cdot 5^k & 2 \cdot 3^k - 5^k \end{bmatrix}$$

A square matrix A is said to be **diagonalizable** if A is similar to a diagonal matrix, that is, if $A = PDP^{-1}$ for some invertible matrix P and some diagonal matrix D.

Impt: when A is diagonalizable, then $A = PDP^{-1}$

Practice Problems 1

PRACTICE PROBLEMS

1. Compute A^8, where $A = \begin{bmatrix} 4 & -3 \\ 2 & -1 \end{bmatrix}$.

SOLUTIONS TO PRACTICE PROBLEMS

1. $\det(A - \lambda I) = \lambda^2 - 3\lambda + 2 = (\lambda - 2)(\lambda - 1)$. The eigenvalues are 2 and 1, and the corresponding eigenvectors are $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Next, form

$$P = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad \text{and} \quad P^{-1} = \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix}$$

Since $A = PDP^{-1}$,

$$A^8 = PD^8P^{-1} = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 2^8 & 0 \\ 0 & 1^8 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 256 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} 766 & -765 \\ 510 & -509 \end{bmatrix}$$

A Real Life Problem: Computing states X[k]

- Initial state of vector X[0]
- Want to know what happens when square matrix A repeatedly move our vector X[0]
- Define X[1] = AX[0]
- Define X[2] = AX[1] = A^2 X[0]
- X[k] = A^(k-1)X[0]

- If we can find high powers of A efficiently, we can know where is the vector X instantly!

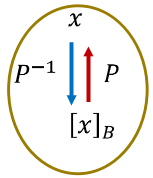
Important point: if we are asked to find x[k=very large number] given $x[n+1] = Ax[n]$ and if A is diagonalizable, it is computationally efficient if we convert $x \to [x]_B$ to work on the problem!

Given A and $x[0]$, and $x[n+1] = Ax[n]$,
we convert problem to B-basis (eigenvectors of A if exists) since its computationally cheaper.
Proof:
$x[1] = Ax[0]$
$x[2] = Ax[1] = A(Ax[0]) = A^2x[0]$
Therefore, $x[k] = A^k x[0]$

if $A = PDP^{-1}$,
then $A^2 = (PDP^{-1})(PDP^{-1}) = PD^2P^{-1}$ and
$A^3 = (PDP^{-1})(PD^2P^{-1}) = PD^3P^{-1}$
Therefore $A^k = PD^kP^{-1}$

$\Rightarrow x[k] = A^k x[0] = PD^kP^{-1}x[0]$
$\Rightarrow x[k] = P\, D^k\, x[0]_B$

Note:
Computing A^k directly is expensive if A is a dense matrix.
Computing D^k is cheap since it is a diagonal matrix.



Change of Basis

$x = P[x]_B$
$P^{-1}x = [x]_B$

Trajectory of Dynamical System

The equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$ determines an infinite collection of equations. Beginning with an initial vector $\mathbf{x}_0$, we have

$$\begin{aligned} \mathbf{x}_1 &= A\mathbf{x}_0 \\ \mathbf{x}_2 &= A\mathbf{x}_1 \\ \mathbf{x}_3 &= A\mathbf{x}_2 \\ &\vdots \end{aligned}$$

The set $\{\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots\}$ is called a trajectory of the system.
Note that $\mathbf{x}_k = A^k\mathbf{x}_0$.

Eigenvalues and eigenvectors provide the key to understanding the long-term behavior, or *evolution*, of a dynamical system described by a difference equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$.

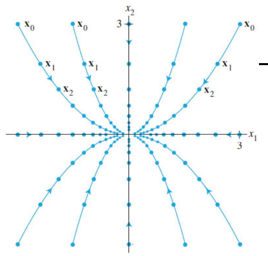


FIGURE 1 The origin as an attractor.

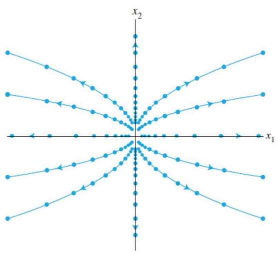


FIGURE 2 The origin as a repeller.

Example 2: when A is a diagonal matrix

Graphical Description of Solutions

When A is 2×2, algebraic calculations can be supplemented by a geometric description of a system's evolution. We can view the equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$ as a description of what happens to an initial point $\mathbf{x}_0$ in $\mathbb{R}^2$ as it is transformed repeatedly by the mapping $\mathbf{x} \mapsto A\mathbf{x}$. The graph of $\mathbf{x}_0, \mathbf{x}_1, \dots$ is called a **trajectory** of the dynamical system.

EXAMPLE 2 Plot several trajectories of the dynamical system $\mathbf{x}_{k+1} = A\mathbf{x}_k$, when

$$A = \begin{bmatrix} .80 & 0 \\ 0 & .64 \end{bmatrix}$$

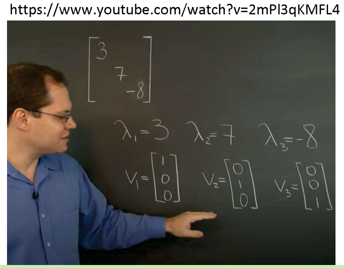
SOLUTION The eigenvalues of A are .8 and .64, with eigenvectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. If $\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$, then

$$\mathbf{x}_k = c_1(.8)^k \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2(.64)^k \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Of course, $\mathbf{x}_k$ tends to $\mathbf{0}$ because $(.8)^k$ and $(.64)^k$ both approach 0 as $k \to \infty$. But *the way* $\mathbf{x}_k$ goes toward $\mathbf{0}$ is interesting. Figure 1 (on page 304) shows the first few terms of several trajectories that begin at points on the boundary of the box with corners at $(\pm 3, \pm 3)$. The points on each trajectory are connected by a thin curve, to make the trajectory easier to see. ■

Lay (4th), p303

https://www.youtube.com/watch?v=2mPl3qKMFL4



Why the standard basis is the eigen vector of a diagonal matrix is easily seen when you pluck these std basis as x into Ax to observe the output.

Why

If $\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$, then

$$\mathbf{x}_k = c_1(.8)^k\begin{bmatrix}1\\0\end{bmatrix} + c_2(.64)^k\begin{bmatrix}0\\1\end{bmatrix}$$

Proof: Lay 4th, pg 278, example 5

If $x_0 = \begin{bmatrix}c_1\\c_2\end{bmatrix}$, then $x_0 = c_1\begin{bmatrix}1\\0\end{bmatrix} + c_2\begin{bmatrix}0\\1\end{bmatrix}$

$$x_0 = c_1 v_1 + c_2 v_2$$

$$x_1 = Ax_0 = A(c_1 v_1 + c_2 v_2)$$
$$= c_1\lambda_1 v_1 + c_2\lambda_2 v_2$$

$$x_2 = Ax_1 = A(c_1\lambda_1 v_1 + c_2\lambda_2 v_2)$$

$$= c_1(\lambda_1)^2\, v_1 + c_2(\lambda_2)^2 v_2$$

Therefore, x_k

$$x_k = c_1(\lambda_1)^k\, v_1 + c_2(\lambda_2)^k v_2$$

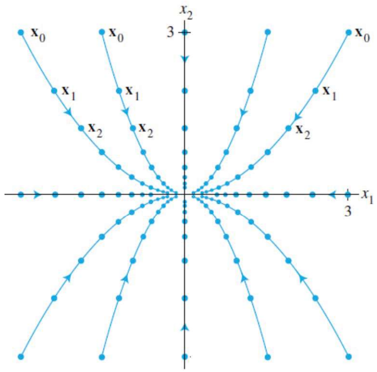


FIGURE 1 The origin as an attractor.

In Example 2, the origin is called an **attractor** of the dynamical system because all trajectories tend toward **0**. This occurs whenever both eigenvalues are less than 1 in magnitude. The direction of greatest attraction is along the line through **0** and the eigenvector $\mathbf{v}_2$ for the eigenvalue of smaller magnitude.

Example 3: repeller at origin

In the next example, both eigenvalues of A are larger than 1 in magnitude, and **0** is called a **repeller** of the dynamical system. All solutions of $\mathbf{x}_{k+1} = A\mathbf{x}_k$ except the (constant) zero solution are unbounded and tend away from the origin.[2]

EXAMPLE 3 Plot several typical solutions of the equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$, where

$$A = \begin{bmatrix}1.44 & 0\\0 & 1.2\end{bmatrix}$$

SOLUTION The eigenvalues of A are 1.44 and 1.2. If $\mathbf{x}_0 = \begin{bmatrix}c_1\\c_2\end{bmatrix}$, then

$$\mathbf{x}_k = c_1(1.44)^k\begin{bmatrix}1\\0\end{bmatrix} + c_2(1.2)^k\begin{bmatrix}0\\1\end{bmatrix}$$

Both terms grow in size, but the first term grows faster. So the direction of greatest repulsion is the line through **0** and the eigenvector for the eigenvalue of larger magnitude. Figure 2 shows several trajectories that begin at points quite close to **0**. ■

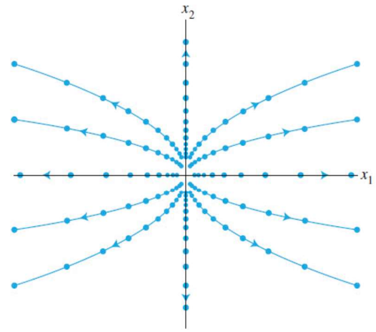


FIGURE 2 The origin as a repeller.

Lay (4th), p305

Example 4: saddle point

EXAMPLE 4 Plot several typical solutions of the equation $\mathbf{y}_{k+1} = D\mathbf{y}_k$, where

$$D = \begin{bmatrix}2.0 & 0\\0 & 0.5\end{bmatrix}$$

(We write D and $\mathbf{y}$ here instead of A and $\mathbf{x}$ because this example will be used later.) Show that a solution $\{\mathbf{y}_k\}$ is unbounded if its initial point is not on the x_2-axis.

SOLUTION The eigenvalues of D are 2 and .5. If $\mathbf{y}_0 = \begin{bmatrix}c_1\\c_2\end{bmatrix}$, then

$$\mathbf{y}_k = c_1 2^k\begin{bmatrix}1\\0\end{bmatrix} + c_2(.5)^k\begin{bmatrix}0\\1\end{bmatrix} \qquad (8)$$

If $\mathbf{y}_0$ is on the x_2-axis, then $c_1 = 0$ and $\mathbf{y}_k \to \mathbf{0}$ as $k \to \infty$. But if $\mathbf{y}_0$ is not on the x_2-axis, then the first term in the sum for $\mathbf{y}_k$ becomes arbitrarily large, and so $\{\mathbf{y}_k\}$ is unbounded. Figure 3 shows ten trajectories that begin near or on the x_2-axis. ■

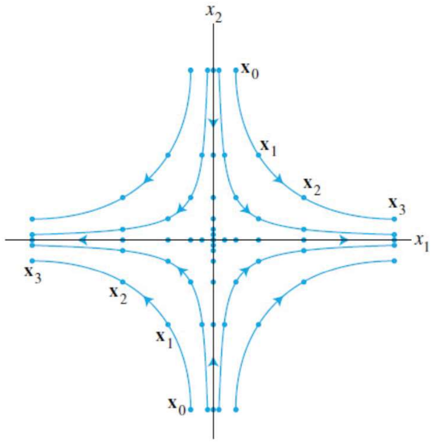


FIGURE 3 The origin as a saddle point.

Lay Example: Dynamical System

EXAMPLE 5 Let $A = \begin{bmatrix}.95 & .03\\.05 & .97\end{bmatrix}$. Analyze the long-term behavior of the dynamical system defined by $\mathbf{x}_{k+1} = A\mathbf{x}_k$ $(k = 0, 1, 2, \ldots)$, with $\mathbf{x}_0 = \begin{bmatrix}.6\\.4\end{bmatrix}$. ← starting pt

SOLUTION The first step is to find the eigenvalues of A and a basis for each eigenspace. The characteristic equation for A is

$$0 = \det\begin{bmatrix}.95-\lambda & .03\\.05 & .97-\lambda\end{bmatrix} = (.95-\lambda)(.97-\lambda) - (.03)(.05)$$
$$= \lambda^2 - 1.92\lambda + .92$$

By the quadratic formula

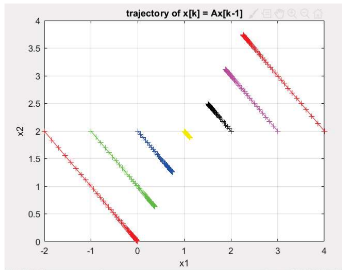
$$\lambda = \frac{1.92 \pm \sqrt{(1.92)^2 - 4(.92)}}{2} = \frac{1.92 \pm \sqrt{.0064}}{2}$$
$$= \frac{1.92 \pm .08}{2} = 1 \quad\text{or}\quad .92$$

It is readily checked that eigenvectors corresponding to $\lambda = 1$ and $\lambda = .92$ are multiples of

$$\mathbf{v}_1 = \begin{bmatrix}3\\5\end{bmatrix} \quad\text{and}\quad \mathbf{v}_2 = \begin{bmatrix}1\\-1\end{bmatrix}$$

respectively.

Lay4thEdition: sec5.2,pg278




```
A  = [0.95 0.03; 0.05 0.97];
N  = 10000

close all;
figure(idxFig);
figID =1;
idxFig = 0;
for  (y = 2:-1:2)
    for (x = -2:1:4)
     x0 = [x y];
     bruteForce_computeAx(A,x0,N,figID,idxFig);
     idxFig=idxFig+1
    end
end
```

```
function bruteForce_computeAx(A,x0,N, figID, idxFig)
colormapStr = 'rgbykm';
x  = x0(:);
hist_x = [];
% brute force way to compute x[k] = Ax[k-1]
for k=0:N
   hist_x = [hist_x; x'];
   x = A*x;
end
tt = mod(idxFig,length(colormapStr))+1;
colorStr = [colormapStr(tt),'-+'];
plot(hist_x(:,1),hist_x(:,2),colorStr)
xlabel('x1'); ylabel('x2');
title('trajectory of x[k] = Ax[k-1]');
grid on; hold on;
end
```

Lay Example: Dynamical System

The next step is to write the given $\mathbf{x}_0$ in terms of $\mathbf{v}_1$ and $\mathbf{v}_2$. This can be done because $\{\mathbf{v}_1, \mathbf{v}_2\}$ is obviously a basis for $\mathbb{R}^2$. (Why?) So there exist weights c_1 and c_2 such that

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = [\mathbf{v}_1 \quad \mathbf{v}_2]\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \tag{3}$$

Very important concept!!!
We express x in terms of the eigen-basis of A!

In fact,

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = [\mathbf{v}_1 \quad \mathbf{v}_2]^{-1}\mathbf{x}_0 = \begin{bmatrix} 3 & 1 \\ 5 & -1 \end{bmatrix}^{-1}\begin{bmatrix} .60 \\ .40 \end{bmatrix}$$

$$= \frac{1}{-8}\begin{bmatrix} -1 & -1 \\ -5 & 3 \end{bmatrix}\begin{bmatrix} .60 \\ .40 \end{bmatrix} = \begin{bmatrix} .125 \\ .225 \end{bmatrix} \tag{4}$$

Because $\mathbf{v}_1$ and $\mathbf{v}_2$ in (3) are eigenvectors of A, with $A\mathbf{v}_1 = \mathbf{v}_1$ and $A\mathbf{v}_2 = .92\mathbf{v}_2$, we easily compute each $\mathbf{x}_k$:

$$\begin{aligned} \mathbf{x}_1 = A\mathbf{x}_0 &= c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2 && \text{Using linearity of } \mathbf{x} \mapsto A\mathbf{x} \\ &= c_1\mathbf{v}_1 + c_2(.92)\mathbf{v}_2 && \mathbf{v}_1 \text{ and } \mathbf{v}_2 \text{ are eigenvectors.} \\ \mathbf{x}_2 = A\mathbf{x}_1 &= c_1A\mathbf{v}_1 + c_2(.92)A\mathbf{v}_2 \\ &= c_1\mathbf{v}_1 + c_2(.92)^2\mathbf{v}_2 \end{aligned}$$

and so on. In general,

$$\mathbf{x}_k = c_1\mathbf{v}_1 + c_2(.92)^k\mathbf{v}_2 \quad (k = 0, 1, 2, \ldots)$$

Using c_1 and c_2 from (4),

$$\mathbf{x}_k = .125\begin{bmatrix} 3 \\ 5 \end{bmatrix} + .225(.92)^k\begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad (k = 0, 1, 2, \ldots) \tag{5}$$

This explicit formula for $\mathbf{x}_k$ gives the solution of the difference equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$. As $k \to \infty$, $(.92)^k$ tends to zero and $\mathbf{x}_k$ tends to $\begin{bmatrix} .375 \\ .625 \end{bmatrix} = .125\mathbf{v}_1$. ■

Dynamical Example: in Tut 8

PRACTICE PROBLEMS

1. The matrix A below has eigenvalues 1, $\frac{2}{3}$, and $\frac{1}{3}$, with corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$:

$$A = \frac{1}{9}\begin{bmatrix} 7 & -2 & 0 \\ -2 & 6 & 2 \\ 0 & 2 & 5 \end{bmatrix}, \quad \mathbf{v}_1 = \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}$$

Find the general solution of the equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$ if $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 11 \\ -2 \end{bmatrix}$.

2. What happens to the sequence $\{\mathbf{x}_k\}$ in Practice Problem 1 as $k \to \infty$?

Dynamical Example

- Since there are 3 eigenvectors from distinct eigenvalues, they form a linearly indep set, and hence is a basis for R^3.
- Impt Steps:

1. Express required vector x0 in terms of the 3 eigenvectors!
2. Note Ax = kx implies A^N(x) = k^N (x)! (k-eigenvalue)

Dynamical Example

SOLUTIONS TO PRACTICE PROBLEMS

1. The first step is to write $\mathbf{x}_0$ as a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. Row reduction of $[\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3 \quad \mathbf{x}_0\,]$ produces the weights $c_1 = 2$, $c_2 = 1$, and $c_3 = 3$, so that

$$\mathbf{x}_0 = 2\mathbf{v}_1 + 1\mathbf{v}_2 + 3\mathbf{v}_3$$

Since the eigenvalues are 1, $\frac{2}{3}$, and $\frac{1}{3}$, the general solution is

$$\begin{aligned} \mathbf{x}_k &= 2 \cdot 1^k\mathbf{v}_1 + 1 \cdot \left(\frac{2}{3}\right)^k\mathbf{v}_2 + 3 \cdot \left(\frac{1}{3}\right)^k\mathbf{v}_3 \\ &= 2\begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix} + \left(\frac{2}{3}\right)^k\begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} + 3 \cdot \left(\frac{1}{3}\right)^k\begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix} \end{aligned} \tag{12}$$

2. As $k \to \infty$, the second and third terms in (12) tend to the zero vector, and

$$\mathbf{x}_k = 2\mathbf{v}_1 + \left(\frac{2}{3}\right)^k\mathbf{v}_2 + 3\left(\frac{1}{3}\right)^k\mathbf{v}_3 \to 2\mathbf{v}_1 = \begin{bmatrix} -4 \\ 4 \\ 2 \end{bmatrix}$$

Notes on Vectors (Refresher)

$\overrightarrow{P_1P_2} = \overrightarrow{OP_2} - \overrightarrow{OP_1}$

↑ Final vector ↑ Terminal point ↑ initial point

Euclidean Space:

$(\mathbb{R}^2 \to x_1, x_2)$ (2-D)

$(\mathbb{R}^3 \to x_1, x_2, x_3)$ (3-D)

To Represent R^3 vector $v \to \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

Norm of a Vector

Magnitude / Length of Vector

$\|v\| = \sqrt{v_1^2 + v_2^2}$ (2D)

$\|v\| = \sqrt{v_1^2 + v_2^2 + v_3^2}$ (3D)

Theorem 3.2.1

If V is a vector in R^n, and k is a scalar,

then: $\|v\| \geq 0$, $\|v\| = 0$ if $v = 0$, $\|kv\| = |k|\,\|v\|$

Unit Length Vector

$u = \frac{1}{\|v\|} v$

Dot Product

$u \cdot v = \|u\|\,\|v\| \cos\theta$

If $u = 0 / v = 0$, then $u \cdot v = 0$

Finding Angle

$\cos\theta = \frac{u \cdot v}{\|u\|\,\|v\|}$ Acute if $u \cdot v > 0$

Obtuse if $u \cdot v < 0$ $\pi/2$ if $u \cdot v = 0$

Properties

$u \cdot v = v \cdot u$

$u \cdot (v + w) = u \cdot v + u \cdot w$

$k(u \cdot v) = (ku) \cdot v$

$v \cdot v \geq 0$ and $v \cdot v = 0$ when $v = 0$

Properties

$0 \cdot v = v \cdot 0 = 0$

$(u + v) \cdot w = u \cdot w + v \cdot w$

$u \cdot (v - w) = u \cdot v - u \cdot w$

$k(u \cdot v) = u \cdot (kv)$

Additional Properties

$Au \cdot v = u \cdot A^T v$

$u \cdot Av = A^T u \cdot v$

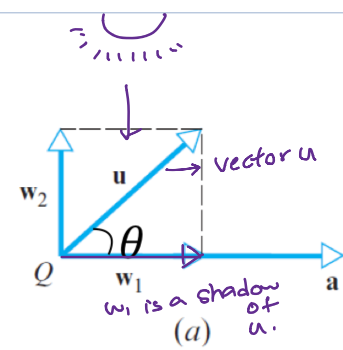
Cauchy-Schwarz Inequality	Triangle Inequality	Parallelogram Equation for Vectors
Theorem 3.2.4	Theorem 3.2.5	Theorem 3.2.6
$-1 \leq \frac{u \cdot v}{\|u\|\|v\|} \leq 1$	$\|u+v\| \leq \|u\| + \|v\|$	$\|u+v\|^2 + \|u-v\|^2 = 2(\|u\|^2 + \|v\|^2)$
$\lvert u \cdot v \rvert \leq \|u\|\|v\|$	$d(u,v) \leq d(u,w) + d(w,v)$	Theorem 3.2.7
		$u \cdot v = \frac{1}{4}\|u+v\|^2 - \frac{1}{4}\|u-v\|^2$

Orthogonality Definition	Orthogonal Lines and Planes	Find the plane through the
Two non-zero vectors	$n \cdot \overrightarrow{P_0 P} = 0$	point $(1,4,9)$ with normal
u and v in R^n are said to	P is an arbitary (x,y,z)	$(2,3,4)$ $\therefore$ $N = (2,3,4)$
be orthogonal if $u \cdot v = 0$	P_0 is a point on line or plane	$P_0(1,4,9)$

Orthogonal Projections

Theorem 3.3.2

If u and v are vectors in R^n, and if $a \neq 0$,
then u can be expressed in exactly one way
in the form $u = w_1 + w_2$, where w_1 is a scalar
multiple of a and w_2 is orthogonal to a.



(a)

Projection of w_1:

$$w_1 = \text{proj}_a u = \frac{u \cdot a}{\|a\|^2} a = \|u\| \cos\theta$$

$$w_2 = u - \text{proj}_a u = u - \frac{u \cdot a}{\|a\|^2} a$$

Pythagoras Theorem

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2$$

Orthogonal sets

A set of vectors $\{u_1 \dots u_p\}$ in R^n is said to be an orthogonal set if each pair of distinct vectors from the set is ortho, that is

$u_i \cdot u_j = 0$ where $i \neq j$

Orthogonal Basis

An orthogonal set of non-zero vectors in R^n is said to be linearly independent, hence a basis for the subspace spanned by S.

Theorem 4 (Proof)

$$0 = 0 \cdot u_1 = (c_1 u_1 + c_2 u_2 + \dots + c_p u_p) \cdot u_1$$

$$= (c_1 u_1) \cdot u_1 + (c_2 u_2) \cdot u_1 + \dots + (c_p u_p) \cdot u_1$$

$$= c_1 (u_1 \cdot u_1) + c_2 (u_1 \cdot u_2) + \dots c_p (u_1 \cdot u_p)$$

$$= c_1 (u_1 \cdot u_1)$$

Since u_1 is orthogonal to $u_2 \dots u_p$. Since u_1 is non-zero, $u_1 \cdot u_1$ is not zero and so $c_1 = 0$, likewise for $c_2 \dots c_p$. Therefore S is linearly independent.

Orthogonal Basis

An orthogonal basis for a subspace W of R^n is a basis for W that is also an orthogonal set.

Theorem 5

Let $\{u_1, \dots u_p\}$ be an orthogonal for a subspace W of R^n.

For each y in w, the weights in the linear combination

$y = c_1 u_1 + \dots + c_p u_p$ are given by

$$c_j = \frac{y \cdot u_j}{u_j \cdot u_j} \qquad (j = 1 \dots p)$$

General Basis:

$y = c_1 v_1 + c_2 v_2 + \dots + c_n v_n$

ON Basis:

$c_i = \langle y, u_i \rangle u_i$

if $\{u_i, u_2, u_n\}$ is ortho and each of them is a unit vector.

Find the orthogonal Basis

$u_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \; u_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \; u_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}$ Orthogonal Basis

Express vector $y = \begin{bmatrix} 6 \\ 1 \\ 8 \end{bmatrix}$ as a linear combination of vectors

Compute $\quad y \cdot u_1 = 11 \quad y \cdot u_2 = -12 \quad y \cdot u_3 = -33$

$\qquad\quad u_1 \cdot u_1 = 11 \quad u_2 \cdot u_2 = 6 \quad u_3 \cdot u_3 = 33/2$

$$= \frac{11}{11} u_1 + \frac{-12}{6} u_2 + \frac{-33}{33/2} u_3$$

$$= u_1 - 2u_2 - 2u_3$$

Orthonormal sets

A set $\{u_1, \ldots, u_p\}$ is an orthormal set if it is an orthogonal set of unit vectors.

If W is the subspace spanned by such a set then $\{u_1, \ldots u_p\}$ is an orthonormal basis for W.

Show a set is an orthonormal basis of R^3

Show dot product of two unique vectors is 0

Show dot product of a vector by itself = 1

E.g.: $U_1 \cdot U_1 = 1$

Theorem 6

An $m \times n$ matrix U has orthonormal columns if only if $U^T U = I$

$u_1^T u_2 = u_2^T u_1 = 0 \qquad u_1^T u_1 = 1$

$u_1^T u_3 = u_3^T u_1 = 0 \qquad u_2^T u_2 = 1$

$u_2^T u_3 = u_3^T u_2 = 0 \qquad u_3^T u_3 = 1$

Orthogonal matrix

A matrix $Q \in R^{m \times n}$ is called orthogonal if $Q^T Q = I_n$, if its columns are orthogonal and have 2-norm one.

If $Q \in R^{m \times n}$ is ortho, then $Q^T Q = I$ implies that $Q^{-1} = Q^T$

If $Q \in R^{m \times n}$ is ortho matrix, then Q^T is an orthogonal matrix.

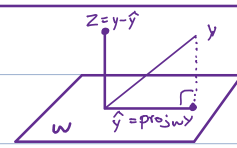
Theorem 7

Let U be an $m \times n$ matrix with orthonormal columns, and let x and y be in R^n. Then

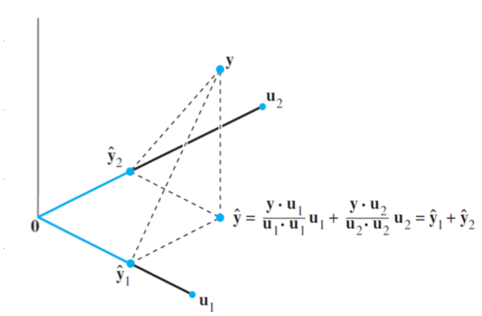
a: $\|Ux\| = \|x\|$

b: $(Ux) \cdot (Uy) = x \cdot y$

c: $(Ux) \cdot (Uy) = 0 \leftrightarrow x \cdot y = 0$



The vector $\hat{y}$ is called the ortho projection of y onto w. Let W be a subspace of R^n can be written uniquely in the form $\underline{y = \hat{y} + z}$ where $\hat{y}$ is in W and z is in $W^{\perp}$. In fact, if $\{u_1, \ldots, u_p\}$ is any orthogonal basis of W, then $\underline{\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \cdots + \frac{y \cdot u_p}{u_p \cdot u_p}}$ and $z = y - \hat{y}$. (Theorem 8)



Let $u_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $u_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, and $y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ Observe that $\{u_1, u_2\}$ is an ortho basis for $W = \text{Span}\{u_1, u_2\}$. Write y as the sum of a vector in W and a vector ortho to W.

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{y \cdot u_2}{u_2 \cdot u_2} u_2 = \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{15}{30}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} \longrightarrow y - \hat{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} = \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

$$y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} + \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

Best Approximation Theorem

Theorem 9

Let W be a subspace of R^n, let y be any vector in R^n, and let $\hat{y}$ be the orthogonal projection of y onto W. Then $\hat{y}$ is the closest point in W to y, in the sense that $\|y-\hat{y}\| < \|y-v\|$ for all v in W distinct from $\hat{y}$.

Properties of Orthogonal Projections

If $\{u_1, \dots, u_p\}$ is an orthogonal basis for W and if y happens to be in W, then the formula for $\text{proj}_W y$ is exactly the same as the representation of y given in Theorem 5 in Section 6.2.

If y is in $W = \text{span}\{u_1, \dots, u_p\}$, then $\text{proj}_W y = y$

Theorem 10

If $\{u_1, \dots, u_p\}$ is an orthonormal basis for a subspace W of R^n, then

$$\text{proj}_W y = (y \cdot u_1)u_1 + (y \cdot u_2)u_2 + \dots + (y \cdot u_p)u_p$$

If $U = [u_1, u_2 \cdots u_p]$, then

$\text{proj}_W y = UU^T y$ for all y in R^n

$U^T U x = I_p x = x$ for all x in R^p

$UU^T y = \text{proj}_W y$ for all y in R^n

Reviewing $Ax = B$

If A can be decomposed into $A = QR$ where, Q is an $m \times n$ matrix with orthonormal columns, R is a Triangular matrix

The problem to find x can then be easily solved by: $Ax = b$

$$QRx = b$$

$$Q^T Q R x = Q^T b$$

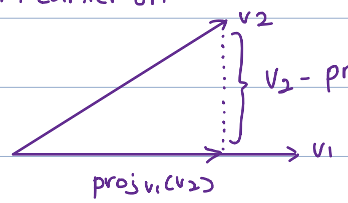
$$Rx = Q^T b$$

Linear Combinations (span)

If $v_1, \dots, v_p$ are in R^n, then the set of all linear combinations of $v_1, \dots, v_p$ is denoted by $\text{Span}\{v_1, \dots, v_p\}$ and is called the subset of R^n spanned by $v_1, \dots, v_p$. That is, $\text{Span}\{v_1, \dots, v_p\}$ is the collection of all vectors that can be written in the form: $c_1v_1 + c_2v_2 + \dots + c_pv_p$

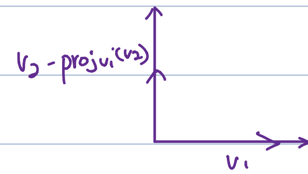
The Gram Schmidt Process → It orthogonalizes a set of vectors

From earlier on



We want the adj and opp

(2 vectors)



The Gram - Schmidt Process

Given a basis $\{x_1, \dots, x_p\}$ for a nonzero subspace W of R^n, define

$$v_1 = x_1$$

$$v_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1$$

$$v_3 = x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} - \frac{x_3 \cdot v_1}{v_2 \cdot v_2} v_2$$

$$\vdots$$

Gram - Schmidt Process for QR Factorisation

Step 1: Split the matrix A into $[x_1, x_2, x_3, \dots]$

Step 2: Perform Gram - schmidt for x_1 till x_p, to obtain v_1 till v_p

Step 3: Combine the v vectors and transpose the matrix

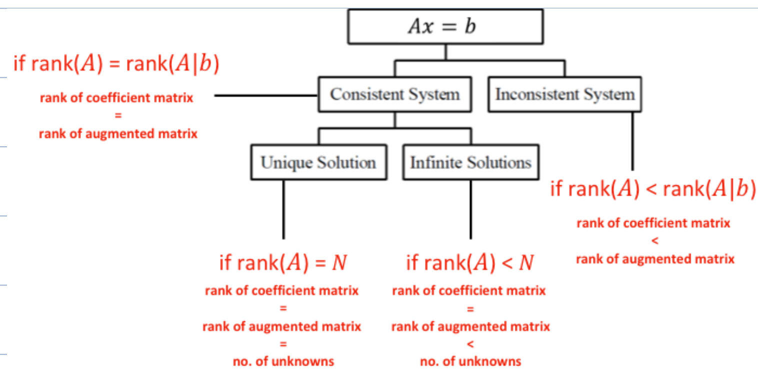
Step 4: $\underline{Q^T A} = R$

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Instead of row to column, we have to use column to column multiplication

Consistency in a System of Equations

Consistent	Inconsistent
b is in column Space of A	b is not in column space of A
b is formed by linear combinations of A's columns.	b is not formed by linear combi's of A's columns
Rank (A) = Rank (A|b), i.e, rank of A is same as that of the Augmented Matrix	Occurs when M >> N (over-determined), there exists more eqn than unknowns
	The rows of A are dependant, but, their corresponding b values are not consistent.
	Rank (A) < Rank (A|b)
	rank of A is less than that of Augmented Matrix

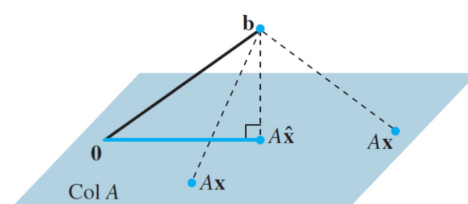


NOTE: Rank (A) is the maximum number of independent rows or columns of A.

You can find number of independent row or columns by:
1. row reduction process
2. rank(A) in MATLAB

Least Square Solution for inconsistent Eqn

Find x such that Ax is as close to b as possible.



Theorem

If A is $m \times n$ and b is in R^m, a least-squares solution of $Ax = b$ is an $\hat{x}$ in R^n such that: $\|b - A\hat{x}\| \le \|b - Ax\|$ for all x in R^n.

$$b - Ax = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix}$$

$$\|b - Ax\|^2 = e_1^2 + e_2^2 + e_3^2 \dots$$

Best Approximation Theorem

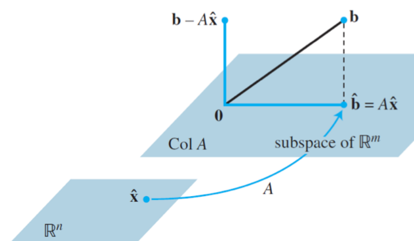
If W is a finite-dimensional subspace of an inner product space V, and if b is a vector in V, then $\text{proj}_W b$ is the best approx to b from W in the sense that:

$\|b - \text{proj}_W b\| < \|b - w\|$ for every vector w in W that is different from $\text{proj}_W b$.

The Normal Equation

$$Ax = b \rightarrow A^T A x = A^T b$$

The set of least-squares solution of $Ax = b$ coincides with the non empty set of solutions of the normal equations $A^T A x = A^T b$.



The Normal Equation Proof

Suppose $\hat{\mathbf{x}}$ satisfies $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. By the Orthogonal Decomposition Theorem in Section 6.3, the projection $\hat{\mathbf{b}}$ has the property that $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to Col A, so $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to each column of A. If $\mathbf{a}_j$ is any column of A, then $\mathbf{a}_j \cdot (\mathbf{b} - A\hat{\mathbf{x}}) = 0$, and $\mathbf{a}_j^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$. Since each $\mathbf{a}_j^T$ is a row of A^T,

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0} \tag{2}$$

(This equation also follows from Theorem 3 in Section 6.1.) Thus

$$A^T\mathbf{b} - A^TA\hat{\mathbf{x}} = \mathbf{0}$$
$$A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$$

These calculations show that each least-squares solution of $A\mathbf{x} = \mathbf{b}$ satisfies the equation

$$A^TA\mathbf{x} = A^T\mathbf{b} \tag{3}$$

The matrix equation (3) represents a system of equations called the **normal equations** for $A\mathbf{x} = \mathbf{b}$. A solution of (3) is often denoted by $\hat{\mathbf{x}}$.

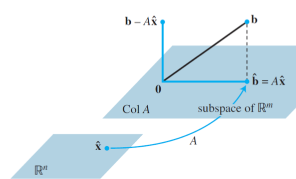


FIGURE 2 The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

THEOREM 14

Let A be an $m \times n$ matrix. The following statements are logically equivalent:

a. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.
b. The columns of A are linearly indpendent.
c. The matrix A^TA is invertible.

When these statements are true, the least-squares solution $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} \tag{4}$$

EXAMPLE 1 Find a least-squares solution of the inconsistent system $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

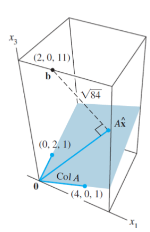
SOLUTION To use normal equations (3), compute:

$$A^TA = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$A^T\mathbf{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

Then the equation $A^TA\mathbf{x} = A^T\mathbf{b}$ becomes

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$



.ay, Linear Algebra and its Applications (4th Edition)

Row operations can be used to solve this system, but since A^TA is invertible and 2×2 it is probably faster to compute

$$(A^TA)^{-1} = \frac{1}{84}\begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$$

and then to solve $A^TA\mathbf{x} = A^T\mathbf{b}$ as

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}$$
$$= \frac{1}{84}\begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}\begin{bmatrix} 19 \\ 11 \end{bmatrix} = \frac{1}{84}\begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

In many calculations, A^TA is invertible, but this is not always the case. The nex

EXAMPLE 2 Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}$$

Note the linear dependency in the rows and columns of A:

- Column 1 = Column 2 + Column 3 + Column 4
- Rows 1 & 2 are same, but their corresponding b values are different (inconsistent)
- Rows 3 & 4 are same, but their corresponding b values are different (inconsistent)
- Rows 5 & 6 are same, but their corresponding b values are different (inconsistent)

SOLUTION Compute

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}}_{A^TA}$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}$$

Note that A^TA is always a square matrix.

The augmented matrix for $A^TA\mathbf{x} = A^T\mathbf{b}$ is

Reduced to

$$\begin{bmatrix} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

A^TA $\quad$ A^Tb

The general solution is $x_1 = 3 - x_4$, $x_2 = -5 + x_4$, $x_3 = -2 + x_4$, and x_4 is free. So the general least-squares solution of $A\mathbf{x} = \mathbf{b}$ has the form

$$\hat{\mathbf{x}} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Applications to Linear Model

Instead of $Ax = b$, we write $X\beta = y$ and refer to X as the design matrix, β as the parameter vector, and y as the observation vector

Least Squares Line

Linear Eqn > $y = \beta_0 + \beta_1 x$

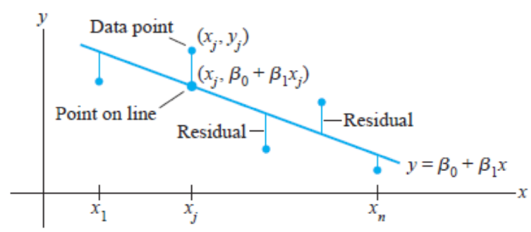


FIGURE 1 Fitting a line to experimental data.

Least Squares Line

To measure how "close" the line is to the data, add the squares of the residuals. The least squares line is the line $y = \beta_0 + \beta_1 x$ that minimizes the sum of the squares residuals.

Predicted y-value		Observed y-value
$\beta_0 + \beta_1 x_1$	=	y_1
$\beta_0 + \beta_1 x_2$	=	y_2
$\vdots$		$\vdots$
$\beta_0 + \beta_1 x_n$	=	y_n

We can write this system as

$$X\beta = \mathbf{y}, \quad \text{where } X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad (1)$$

The square of the distance between the vectors $X\beta$ and $\mathbf{y}$ is precisely the sum of the squares of the residuals. The β that minimizes this sum also minimizes the distance between $X\beta$ and $\mathbf{y}$. *Computing the least-squares solution of $X\beta = \mathbf{y}$ is equivalent to finding the β that determines the least-squares line in Figure 1.*

EXAMPLE 1 Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points (2, 1), (5, 2), (7, 3), and (8, 3).

SOLUTION Use the x-coordinates of the data to build the design matrix X in (1) and the y-coordinates to build the observation vector $\mathbf{y}$:

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$$

For the least-squares solution of $X\beta = \mathbf{y}$, obtain the normal equations (with the new notation):

$$X^T X\beta = X^T\mathbf{y}$$

That is, compute

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

The normal equations are

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

Hence

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 142 & -22 \\ -22 & 4 \end{bmatrix} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 24 \\ 30 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$$

Thus the least-squares line has the equation

$$y = \frac{2}{7} + \frac{5}{14}x$$

Curve fitting

In some applications, it is necessary to fit data points with something other than a straight line. In the examples that follow, the matrix equation is still $X\beta = \mathbf{y}$, but the specific form of X changes from one problem to the next. Statisticians usually introduce a **residual vector** ϵ, defined by $\epsilon = \mathbf{y} - X\beta$, and write

$$\mathbf{y} = X\beta + \epsilon$$

Any equation of this form is referred to as a **linear model**. Once X and $\mathbf{y}$ are determined, the goal is to minimize the length of ϵ, which amounts to finding a least-squares solution of $X\beta = \mathbf{y}$. In each case, the least-squares solution $\hat{\beta}$ is a solution of the normal equations

$$X^T X\beta = X^T\mathbf{y}$$

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 \quad (3)$$

Describe the linear model that produces a "least-squares fit" of the data by equation (3).

SOLUTION Equation (3) describes the ideal relationship. Suppose the actual values of the parameters are $\beta_0, \beta_1, \beta_2$. Then the coordinates of the first data point (x_1, y_1) satisfy an equation of the form

$$y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon_1$$

where ϵ_1 is the residual error between the observed value y_1 and the predicted y-value $\beta_0 + \beta_1 x_1 + \beta_2 x_1^2$. Each data point determines a similar equation:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 + \epsilon_2 \\ &\vdots \qquad\qquad \vdots \\ y_n &= \beta_0 + \beta_1 x_n + \beta_2 x_n^2 + \epsilon_n \end{aligned}$$

It is a simple matter to write this system of equations in the form $\mathbf{y} = X\beta + \epsilon$. To find X, inspect the first few rows of the system and look for the pattern.

$$\underset{\mathbf{y}}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}} = \underset{X}{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}} \underset{\beta}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}} + \underset{\epsilon}{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}}$$

■

4.3 B Find the parabola $C + Dt + Et^2$ that comes closest (least squares error) to the values $\boldsymbol{b} = (0, 0, 1, 0, 0)$ at the times $t = -2, -1, 0, 1, 2$. First write down the five equations $A\boldsymbol{x} = \boldsymbol{b}$ in three unknowns $\boldsymbol{x} = (C, D, E)$ for a parabola to go through the five points. No solution because no such parabola exists. Solve $A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$.

$$\begin{aligned} C + D\,(-2) + E\,(-2)^2 &= 0 \\ C + D\,(-1) + E\,(-1)^2 &= 0 \\ C + D\ \ (0) + E\ \ (0)^2 &= 1 \\ C + D\ \ (1) + E\ \ (1)^2 &= 0 \\ C + D\ \ (2) + E\ \ (2)^2 &= 0 \end{aligned} \qquad A = \begin{bmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \qquad A^T A = \begin{bmatrix} 5 & \mathbf{0} & 10 \\ \mathbf{0} & 10 & \mathbf{0} \\ 10 & \mathbf{0} & 34 \end{bmatrix}$$

Those zeros in $A^T A$ mean that column 2 of A is orthogonal to columns 1 and 3. We see this directly in A (the times $-2, -1, 0, 1, 2$ are symmetric). The best C, D, E in the parabola $C + Dt + Et^2$ come from $A^T A\hat{x} = A^T\boldsymbol{b}$, and D is uncoupled:

$$\begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{leads to} \quad \begin{aligned} C &= 34/70 \\ D &= 0 \ \text{ as predicted} \\ E &= -10/70 \end{aligned}$$