DM554

Linear and Integer Programming

Lecture 9
Diagonalization

Marco Chiarandini

Department of Mathematics & Computer Science
University of Southern Denmark

Outline

1. More on Coordinate Change
2. Diagonalization
3. Applications

Resume

- Linear transformations and proofs that a given mapping is linear
- range and null space, and rank and nullity of a transformation, rank-nullity theorem
- two-way relationship between matrices and linear transformations
- change from standard to arbitrary basis
- change of basis from B to B'

Outline

1. More on Coordinate Change

2. Diagonalization

3. Applications

Change of Basis for a Lin. Transf.

We saw how to find A for a transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ using standard basis in both $\mathbb{R}^n$ and $\mathbb{R}^m$. Now: is there a matrix that represents T wrt two arbitrary bases B and B'?

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and $B' = \{\mathbf{v}'_1, \mathbf{v}'_2, \ldots, \mathbf{v}'_n\}$ be bases of $\mathbb{R}^n$ and $\mathbb{R}^m$. Then for all $\mathbf{x} \in \mathbb{R}^n$, $\quad [T(\mathbf{x})]_{B'} = M[\mathbf{x}]_B$
where $M = A_{[B,B']}$ is the $m \times n$ matrix with the ith column equal to $[T(\mathbf{v}_i)]_{B'}$, the coordinate vector of $T(\mathbf{v}_i)$ wrt the basis B'.

Proof:

change B to standard $\qquad \mathbf{x} = P_B^{n\times n}[\mathbf{x}]_B \quad \forall \mathbf{x} \in \mathbb{R}^n$

$\downarrow$

perform linear transformation in standard coordinates $\qquad T(\mathbf{x}) = A\mathbf{x} = AP_B^{n\times n}[\mathbf{x}]_B$

$\downarrow$

change to basis B' $\qquad [\mathbf{u}]_{B'} = (P_{B'}^{m\times m})^{-1}\mathbf{u} \quad \forall \mathbf{u} \in \mathbb{R}^m$

$$[T(\mathbf{x})]_{B'} = (P_{B'}^{m\times m})^{-1}AP_B^{n\times n}[\mathbf{x}]_B$$

$$M = (P_{B'}^{m\times m})^{-1}AP_B^{n\times n}$$

How is M done?

- $P_B = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_n]$
- $AP_B = A[\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_n] = [A\mathbf{v}_1 \ A\mathbf{v}_2 \ \dots \ A\mathbf{v}_n]$
- $A\mathbf{v}_i = T(\mathbf{v}_i)$: $AP_B = [T(\mathbf{v}_1) \ T(\mathbf{v}_2) \ \dots \ T(\mathbf{v}_n)]$
- $M = P_{B'}^{-1} AP_B = P_{B'}^{-1} = [P_{B'}^{-1} T(\mathbf{v}_1) \ P_{B'}^{-1} T(\mathbf{v}_2) \ \dots \ P_{B'}^{-1} T(\mathbf{v}_n)]$
- $M = [[T(\mathbf{v}_1)]_{B'} \ [T(\mathbf{v}_2)]_{B'} \ \dots \ [T(\mathbf{v}_n)]_{B'}]$

Hence, if we change the basis from the standard basis of $\mathbb{R}^n$ and $\mathbb{R}^m$ the matrix representation of T changes

Similarity

Particular case $m = n$:

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation and $B = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be a basis $\mathbb{R}^n$. Let A be the matrix corresponding to T in standard coordinates: $T(\mathbf{x}) = A\mathbf{x}$. Let

$$P = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}$$

be the matrix whose columns are the vectors of B. Then for all $\mathbf{x} \in \mathbb{R}^n$,

$$[T(\mathbf{x})]_B = P^{-1}AP[\mathbf{x}]_B$$

Or, the matrix $A_{[B,B]} = P^{-1}AP$ performs the same linear transformation as the matrix A but expressed it in terms of the basis B.

Similarity

Definition

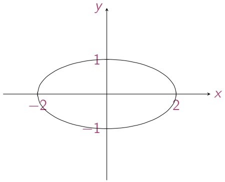
A square matrix C is similar (represent the same linear transformation) to the matrix A if there is an invertible matrix P such that

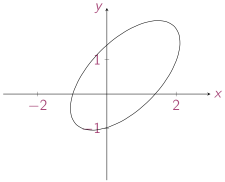
$$C = P^{-1}AP.$$

Similarity defines an equivalence relation:

- (reflexive) a matrix A is similar to itself
- (symmetric) if C is similar to A, then A is similar to C
  $C = P^{-1}AP, \quad A = Q^{-1}CQ, \quad Q = P^{-1}$
- (transitive) if D is similar to C, and C to A, then D is similar to A

Example





- $x^2 + y^2 = 1$ circle in standard form
- $x^2 + 4y^2 = 4$ ellipse in standard form
- $5x^2 + 5y^2 - 6xy = 2$??? Try rotating $\pi/4$ anticlockwise

$$A_T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = P$$

$$\mathbf{v} = P[\mathbf{v}]_B \iff \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}$$

$$X^2 + 4Y^2 = 1$$

Example

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$:

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x + 3y \\ -x + 5y \end{bmatrix}$$

What is its effect on the xy-plane?
Let's change the basis to

$$B = \{\mathbf{v}_1, \mathbf{v}_2\} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}$$

Find the matrix of T in this basis:

- $C = P^{-1}AP$, A matrix of T in standard basis, P is transition matrix from B to standard

$$C = P^{-1}AP = \frac{1}{2}\begin{bmatrix} -1 & 3 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ -1 & 5 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$$

Example (cntd)

- the B coordinates of the B basis vectors are

$$[\mathbf{v}_1]_B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}_B, \quad [\mathbf{v}_2]_B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}_B$$

- so in B coordinates T is a stretch in the direction $\mathbf{v}_1$ by 4 and in dir. $\mathbf{v}_2$ by 2:

$$[T(\mathbf{v}_1)]_B = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}_B = \begin{bmatrix} 4 \\ 0 \end{bmatrix}_B = 4[\mathbf{v}_1]_B$$

- The effect of T is however the same no matter what basis, only the matrices change! So also in the standard coordinates we must have:

$$A\mathbf{v}_1 = 4\mathbf{v}_1 \qquad A\mathbf{v}_2 = 2\mathbf{v}_2$$

Resume

- Matrix representation of a transformation with respect to two given basis
- Similarity of square matrices

Outline

Eigenvalues and Eigenvectors

(All matrices from now on are square $n \times n$ matrices and all vectors in $\mathbb{R}^n$)

Definition

Let A be a square matrix.

- The number λ is said to be an eigenvalue of A if for some non-zero vector $\mathbf{x}$,

$$A\mathbf{x} = \lambda\mathbf{x}$$

- Any non-zero vector $\mathbf{x}$ for which this equation holds is called eigenvector for eigenvalue λ or
eigenvector of A corresponding to eigenvalue λ

Finding Eigenvalues

- Determine solutions to the matrix equation $A\mathbf{x} = \lambda\mathbf{x}$
- Let's put it in standard form, using $\lambda\mathbf{x} = \lambda I\mathbf{x}$:

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

- $B\mathbf{x} = \mathbf{0}$ has solutions other than $\mathbf{x} = \mathbf{0}$ precisely when $\det(B) = 0$.
- hence we want $\det(A - \lambda I) = 0$:

Definition (Charachterisitc polynomial)

The polynomial $|A - \lambda I|$ is called the characteristic polynomial of A, and the equation $|A - \lambda I| = 0$ is called the characteristic equation of A.

Example

$$A = \begin{bmatrix} 7 & -15 \\ 2 & -4 \end{bmatrix}$$

$$A - \lambda I = \begin{bmatrix} 7 & -15 \\ 2 & -4 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 7-\lambda & -15 \\ 2 & -4-\lambda \end{bmatrix}$$

The characteristic polynomial is

$$\begin{aligned} |A - \lambda I| &= \begin{vmatrix} 7-\lambda & -15 \\ 2 & -4-\lambda \end{vmatrix} \\ &= (7-\lambda)(-4-\lambda) + 30 \\ &= \lambda^2 - 3\lambda + 2 \end{aligned}$$

The characteristic equation is

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

hence 1 and 2 are the only eigenvalues of A

Finding Eigenvectors

- Find non-trivial solution to $(A - \lambda I)\mathbf{x} = \mathbf{0}$ corresponding to λ
- zero vectors are not eigenvectors!

Example

$$A = \begin{bmatrix} 7 & -15 \\ 2 & -4 \end{bmatrix}$$

Eigenvector for $\lambda = 1$:

$$A - I = \begin{bmatrix} 6 & -15 \\ 2 & -5 \end{bmatrix} \to \overset{RREF}{\cdots} \to \begin{bmatrix} 1 & -\frac{5}{2} \\ 0 & 0 \end{bmatrix} \qquad \mathbf{v} = t \begin{bmatrix} 5 \\ 2 \end{bmatrix},\ t \in \mathbb{R}$$

Eigenvector for $\lambda = 2$:

$$A - 2I = \begin{bmatrix} 5 & -15 \\ 2 & -6 \end{bmatrix} \to \overset{RREF}{\cdots} \to \begin{bmatrix} 1 & -3 \\ 0 & 0 \end{bmatrix} \qquad \mathbf{v} = t \begin{bmatrix} 3 \\ 1 \end{bmatrix},\ t \in \mathbb{R}$$

Example

$$A = \begin{bmatrix} 4 & 0 & 4 \\ 0 & 4 & 4 \\ 4 & 4 & 8 \end{bmatrix}$$

The characteristic equation is

$$\begin{aligned} |A - \lambda I| &= \begin{vmatrix} 4-\lambda & 0 & 4 \\ 0 & 4-\lambda & 4 \\ 4 & 4 & 8-\lambda \end{vmatrix} \\ &= (4-\lambda)((-4-\lambda)(8-\lambda) - 16) + 4(-4(4-\lambda)) \\ &= (4-\lambda)((-4-\lambda)(8-\lambda) - 16) - 16(4-\lambda) \\ &= (4-\lambda)((-4-\lambda)(8-\lambda) - 16 - 16) \\ &= (4-\lambda)\lambda(\lambda - 12) \end{aligned}$$

hence the eigenvalues are $4, 0, 12$.
Eigenvector for $\lambda = 4$, solve $(A - 4I)\mathbf{x} = \mathbf{0}$:

$$A - 4I = \begin{bmatrix} 4-4 & 0 & 4 \\ 0 & 4-4 & 4 \\ 4 & 4 & 8-4 \end{bmatrix} \rightarrow \overset{RREF}{\cdots} \rightarrow \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{v} = t \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \; t \in \mathbb{R}$$

Example

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

The characteristic equation is

$$\begin{aligned} |A - \lambda I| &= \begin{vmatrix} -3-\lambda & -1 & -2 \\ 1 & -1-\lambda & 1 \\ 1 & 1 & -\lambda \end{vmatrix} \\ &= (-3-\lambda)(\lambda^2+\lambda-1) + (-\lambda-1) - 2(2+\lambda) \\ &= -(\lambda^3 + 4\lambda^2 + 5\lambda + 2) \end{aligned}$$

if we discover that -1 is a solution then $(\lambda + 1)$ is a factor of the polynomial:

$$-(\lambda+1)(a\lambda^2 + b\lambda + c)$$

from which we can find $a = 1, c = 2, b = 3$ and

$$-(\lambda+1)(\lambda+2)(\lambda+1) = -(\lambda+1)^2(\lambda+2)$$

the eigenvalue -1 has multiplicity 2

Eigenspaces

- The set of eigenvectors corresponding to the eigenvalue λ together with the zero vector $\mathbf{0}$, is a subspace of $\mathbb{R}^n$.
  because it corresponds with null space $N(A - \lambda I)$

Definition (Eigenspace)

If A is an $n \times n$ matrix and λ is an eigenvalue of A, then the eigenspace of the eigenvalue λ is the nullspace $N(A - \lambda I)$ of $\mathbb{R}^n$.

- the set $S = \{\mathbf{x} \mid A\mathbf{x} = \lambda\mathbf{x}\}$ is always a subspace but only if λ is an eigenvalue then $\dim(S) \geq 1$.

Eigenvalues and the Matrix

Links between eigenvalues and properties of the matrix

- let A be an $n \times n$ matrix, then the characteristic polynomial has degree n:

$$p(\lambda) = |A - \lambda I| = (-1)^n(\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0)$$

- in terms of eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ the characteristic polynomial is:

$$p(\lambda) = |A - \lambda I| = (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$$

Theorem

The determinant of an $n \times n$ matrix A is equal to the product of its eigenvalues.

Proof: if $\lambda = 0$ in the first point above, then

$$p(0) = |A| = (-1)^n a_0 = (-1)^n(-1)^n\lambda_1\lambda_2\ldots\lambda_n = \lambda_1\lambda_2\ldots\lambda_n$$

- The trace of a square matrix A is the sum of the entries on its main diagonal.

Theorem

The trace of an $n \times n$ matrix is equal to the sum of its eigenvalues.

Proof:

$$\begin{aligned} |A - \lambda I| &= (-1)^n(\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0) \\ &= (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n) \end{aligned}$$

the proof follows by comparing the coefficients of $(-\lambda)^{n-1}$

Diagonalization

Recall: Square matrices are similar if there is an invertible matrix P such that $P^{-1}AP = M$.

Definition (Diagonalizable matrix)

The matrix A is diagonalizable if it is similar to a diagonal matrix; that is, if there is a diagonal matrix D and an invertible matrix P such that $P^{-1}AP = D$

Example

$$A = \begin{bmatrix} 7 & -15 \\ 2 & -4 \end{bmatrix}$$

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

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

How was such a matrix P found?

When a matrix is diagonalizable?

General Method

- Let's assume A is diagonalizable, then $P^{-1}AP = D$ where

$$D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

- $AP = PD$

$$AP = A \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{v}_1 & \cdots & A\mathbf{v}_n \end{bmatrix}$$

$$PD = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{v}_1 & \cdots & \lambda_n\mathbf{v}_n \end{bmatrix}$$

- Hence: $A\mathbf{v}_1 = \lambda_1\mathbf{v}_1, \quad A\mathbf{v}_2 = \lambda_2\mathbf{v}_2, \quad \cdots \quad A\mathbf{v}_n = \lambda_n\mathbf{v}_n$

- since P^{-1} exists then none of the above $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ has $\mathbf{0}$ as a solution or else P would have a zero column.
- this is equivalent to λ_i and $\mathbf{v}_i$ are eigenvalues and eigenvectors and that they are linearly independent.
- the converse is also true: P^{-1} is invertible and $A\mathbf{v} = \lambda\mathbf{v}$ implies that

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

Theorem

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

Theorem

An $n \times n$ matrix A is diagonalizable if and only if there is a basis of $\mathbb{R}^n$ consisting only of eigenvectors of A.

Example

$$A = \begin{bmatrix} 7 & -15 \\ 2 & -4 \end{bmatrix}$$

and 1 and 2 are the eigenvalues with eigenvectors:

$$\mathbf{v}_1 = \begin{bmatrix} 5 \\ 2 \end{bmatrix} \qquad \mathbf{v}_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 2 & 1 \end{bmatrix}$$

Example

$$A = \begin{bmatrix} 4 & 0 & 4 \\ 0 & 4 & 4 \\ 4 & 4 & 8 \end{bmatrix}$$

has eigenvalues $0, 4, 12$ and corresponding eigenvectors:

$$\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

$$P = \begin{bmatrix} -1 & -1 & 1 \\ 1 & -1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \qquad D = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 12 \end{bmatrix}$$

We can choose any order, provided we are consistent:

$$P = \begin{bmatrix} -1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix} \qquad D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 12 \end{bmatrix}$$

Geometrical Interpretation

- Let's look at A as the matrix representing a linear transformation $T = T_A$ in standard coordinates, ie, $T(\mathbf{x}) = A\mathbf{x}$.

- let's assume A has a set of linearly independent vectors $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ corresponding to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then B is a basis of $\mathbb{R}^n$.

- what is the matrix representing T wrt the basis B?

  $$A_{[B,B]} = P^{-1}AP$$

  where $P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$ (check earlier theorem today)

- hence, the matrices A and $A_{[B,B]}$ are similar, they represent the same linear transformation:
  - A in the standard basis
  - $A_{[B,B]}$ in the basis B of eigenvectors of A

- $A_{[B,B]} = \begin{bmatrix} [T(\mathbf{v}_1)]_B & [T(\mathbf{v}_2)]_B & \cdots & [T(\mathbf{v}_n)]_B \end{bmatrix} \rightsquigarrow$ for those vectors in particular $T(\mathbf{v}_i) = A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ hence diagonal matrix $\rightsquigarrow A_{[B,B]} = D$

- What does this tell us about the linear transformation T_A?

$$\text{For any } \mathbf{x} \in \mathbb{R}^n \qquad [\mathbf{x}]_B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}_B$$

  its image in T is easy to calculate in B coordinates:

$$[T(\mathbf{x})]_B = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}_B = \begin{bmatrix} \lambda_1 b_1 \\ \lambda_2 b_2 \\ \vdots \\ \lambda_n b_n \end{bmatrix}_B$$

- it is a stretch in the direction of the eigenvector $\mathbf{v}_i$ by a factor λ_i!

- the line $\mathbf{x} = t\mathbf{v}_i$, $t \in \mathbb{R}$ is fixed by the linear transformation T in the sense that every point on the line is stretched to another point on the same line.

Similar Matrices

Geometric interpretation

- Let A and $B = P^{-1}AP$, ie, be similar.
- geometrically: T_A is a linear transformation in standard coordinates T_B is the same linear transformation T in coordinates wrt the basis given by the columns of P.
- we have seen that T has the intrinsic property of fixed lines and stretches. This property does not depend on the coordinate system used to express the vectors. Hence:

Theorem

Similar matrices have the same eigenvalues, and the same corresponding eigenvectors expressed in coordinates with respect to different bases.

Algebraically:

- A and B have same polynomial and hence eigenvalues

$$\begin{aligned} |B - \lambda I| &= |P^{-1}AP - \lambda I| = |P^{-1}AP - \lambda P^{-1}IP| \\ &= |P^{-1}(A - \lambda I)P| = |P^{-1}||A - \lambda I||P| \\ &= |A - \lambda I| \end{aligned}$$

- P transition matrix from the basis S to the standard coords to coords

$$\mathbf{v} = P[\mathbf{v}]_S \qquad [\mathbf{v}]_S = P^{-1}\mathbf{v}$$

- Using $A\mathbf{v} = \lambda\mathbf{v}$:

$$\begin{aligned} B[\mathbf{v}]_S &= P^{-1}AP[\mathbf{v}]_S \\ &= P^{-1}A\mathbf{v} \\ &= P^{-1}\lambda\mathbf{v} \\ &= \lambda P^{-1}\mathbf{v} \\ &= \lambda[\mathbf{v}]_S \end{aligned}$$

hence $[\mathbf{v}]_S$ is eigenvector of B corresponding to eigenvalue λ

Diagonalizable matrices

Example

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

has characteristic polynomial $\lambda^2 - 6\lambda + 9 = (\lambda - 3)^2$.
The eigenvectors are:

$$\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\mathbf{v} = [-1, 1]^T$$

hence any two eigenvectors are scalar multiple of each others and are linearly dependent.

The matrix A is therefore not diagonalizable.

Example

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

has characteristic equation $\lambda^2 + 1$ and hence it has no real eigenvalues.

Theorem

If an $n \times n$ matrix A has n different eigenvalues then (it has a set of n linearly independent eigenvectors) is diagonalizable.

- Proof by contradiction
- n lin indep. is necessary condition but n different eigenvalues not.

Example

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

the characteristic polynomial is $-(\lambda - 2)^2(\lambda - 4)$. Hence 2 has multiplicity 2. Can we find two corresponding linearly independent vectors?

Example (cntd)

$$(A-2I)=\begin{bmatrix}1&-1&1\\0&0&0\\1&-1&1\end{bmatrix}\to\overset{RREF}{\cdots}\to\begin{bmatrix}1&-1&1\\0&0&0\\0&0&0\end{bmatrix}$$

$$\mathbf{x}=s\begin{bmatrix}1\\1\\0\end{bmatrix}+t\begin{bmatrix}-1\\0\\1\end{bmatrix}=s\mathbf{v}_1+t\mathbf{v}_2\quad s,t\in\mathbb{R}$$

the two vectors are lin. indep.

$$(A-4I)=\begin{bmatrix}-1&-1&1\\0&-2&0\\1&-1&-1\end{bmatrix}\to\overset{RREF}{\cdots}\to\begin{bmatrix}1&0&-1\\0&1&0\\0&0&0\end{bmatrix}\qquad\mathbf{v}_3=\begin{bmatrix}1\\0\\1\end{bmatrix}$$

$$P=\begin{bmatrix}1&1&-1\\0&1&0\\1&0&1\end{bmatrix}\qquad P^{-1}AP=\begin{bmatrix}4&0&0\\0&2&0\\0&0&2\end{bmatrix}$$

Example

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

Eigenvalue $\lambda_1 = -1$ has multiplicity 2; $\lambda_2 = -2$.

$$(A + I) = \begin{bmatrix} -2 & -1 & -2 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \rightarrow \overset{RREF}{\cdots} \rightarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The rank is 2.
The null space $(A + I)$ therefore has dimension 1 (rank-nullity theorem).
We find only one linearly independent vector: $\mathbf{x} = [-1, 0, 1]^T$.
Hence the matrix A cannot be diagonalized.

Multiplicity

Definition (Algebraic and geometric multiplicity)

An eigenvalue λ_0 of a matrix A has

- algebraic multiplicity k if k is the largest integer such that $(\lambda - \lambda_0)^k$ is a factor of the characteristic polynomial
- geometric multiplicity k if k is the dimension of the eigenspace of λ_0, ie, $\dim(N(A - \lambda_0 I))$

Theorem

For any eigenvalue of a square matrix, the geometric multiplicity is no more than the algebraic multiplicity

Theorem

A matrix is diagonalizable if and only if all its eigenvalues are real numbers and, for each eigenvalue, its geometric multiplicity equals the algebraic multiplicity.

Summary

- Characteristic polynomial and characteristic equation of a matrix
- eigenvalues, eigenvectors, diagonalization
- finding eigenvalues and eigenvectors
- eigenspace
- eigenvalues are related to determinant and trace of a matrix
- diagonalize a diagonalizable matrix
- conditions for digonalizability
- diagonalization as a change of basis, similarity
- geometric effect of linear transformation via diagonalization

Outline

1. More on Coordinate Change

2. Diagonalization

3. Applications

Uses of Diagonalization

- find powers of matrices
- solving systems of simultaneous linear difference equations
- Markov chains
- systems of differential equations

Powers of Matrices

$$A^n = \underbrace{AAA\cdots A}_{n \text{ times}}$$

If we can write: $P^{-1}AP = D$ then $A = PDP^{-1}$

$$\begin{aligned}
A^n &= \underbrace{AAA\cdots A}_{n \text{ times}} \\
&= \underbrace{(PDP^{-1})(PDP^{-1})(PDP^{-1})\cdots(PDP^{-1})}_{n \text{ times}} \\
&= PD(P^{-1}P)D(P^{-1}P)D(P^{-1}P)\cdots DP^{-1} \\
&= P\,\underbrace{DDD\cdots D}_{n \text{ times}}\,P^{-1} \\
&= PD^nP^{-1}
\end{aligned}$$

then closed formula to calculate the power of a matrix.

Difference equations

- A difference equation is an equation linking terms of a sequence to previous terms, eg:

  $$x_{t+1} = 5x_t - 1$$

  is a first order difference equation.

- a first order difference equation can be fully determined if we know the first term of the sequence (initial condition)

- a solution is an expression of the terms $\mathbf{x}_t$

  $$x_{t+1} = ax_t \implies x_t = a^t x_0$$

System of Difference equations

Suppose the sequences x_t and y_t are related as follows:
$x_0 = 1, y_0 = 1$ for $t \geq 0$

$$\begin{aligned} x_{t+1} &= 7x_t - 15y_t \\ y_{t+1} &= 2x_t - 4y_t \end{aligned}$$

Coupled system of difference equations.

Let

$$\mathbf{x}_t = \begin{bmatrix} x_t \\ y_t \end{bmatrix}$$

then $\mathbf{x}_{t+1} = A\mathbf{x}_t$ and $\mathbf{0} = [1, 1]^T$ and

$$A = \begin{bmatrix} 7 & -15 \\ 2 & -4 \end{bmatrix}$$

Then:

$$\begin{aligned} \mathbf{x}_1 &= A\mathbf{x}_0 \\ \mathbf{x}_2 &= A\mathbf{x}_1 = A(A\mathbf{x}_0) = A^2\mathbf{x}_0 \\ \mathbf{x}_3 &= A\mathbf{x}_2 = A(A^2\mathbf{x}_0) = A^3\mathbf{x}_0 \\ &\vdots \\ \mathbf{x}_t &= A^t\mathbf{x}_0 \end{aligned}$$

Markov Chains

- Suppose two supermarkets compete for customers in a region with 20000 shoppers.
- Assume no shopper goes to both supermarkets in a week.
- The table gives the probability that a shopper will change from one to another supermarket:

	From A	From B	From none
To A	0.70	0.15	0.30
To B	0.20	0.80	0.20
To none	0.10	0.05	0.50

  (note that probabilities in the columns add up to 1)
- Suppose that at the end of week 0 it is known that 10000 went to A, 8000 to B and 2000 to none.
- Can we predict the number of shoppers at each supermarket in any future week t? And the long-term distribution?

Formulation as a system of difference equations:

- Let $\mathbf{x}_t$ be the percentage of shoppers going in the two supermarkets or none
- then we have the difference equation:

$$\mathbf{x}_t = A\mathbf{x}_{t-1}$$

$$A = \begin{bmatrix} 0.70 & 0.15 & 0.30 \\ 0.20 & 0.80 & 0.20 \\ 0.10 & 0.05 & 0.50 \end{bmatrix}, \qquad \mathbf{x}_t = \begin{bmatrix} x_t & y_t & z_t \end{bmatrix}$$

- a Markov chain (or process) is a closed system of a fixed population distributed into n diffrerent states, transitioning between the states during specific time intervals.
- The transition probabilities are known in a transition matrix A (coefficients all non-negative + sum of entries in the columns is 1)
- state vector $\mathbf{x}_t$, entries sum to 1.

- A solution is given by (assuming A is diagonalizable):

$$\mathbf{x}_t = A^t \mathbf{x}_0 = (PD^t P^{-1})\mathbf{x}_0$$

- let $\mathbf{x}_0 = P\mathbf{z}_0$ and $\mathbf{z}_0 = P^{-1}\mathbf{x}_0 = \begin{bmatrix} b_1 & b_2 \cdots & b_n \end{bmatrix}^T$ be the representation of $\mathbf{x}_0$ in the basis of eigenvectors, then:

$$\mathbf{x}_t = PD^t P^{-1}\mathbf{x}_0 = b_1\lambda_1^t \mathbf{v}_1 + b_2\lambda_2^t \mathbf{v}_2 + \cdots + b_n\lambda_n^t \mathbf{v}_n$$

- $\mathbf{x}_t = b_1(1)^t \mathbf{v}_1 + b_2(0.6)^t \mathbf{v}_2 + \cdots + b_n(0.4)^t \mathbf{v}_n$

- $\lim_{t\to\infty} 1^t = 1, \quad \lim_{t\to\infty} 0.6^t = 0$ hence the long-term distribution is

$$\mathbf{q} = b_1\mathbf{v}_1 = 0.125 \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.500 \\ 0.125 \end{bmatrix}$$

- Th.: if A is the transition matrix of a regular Markov chain, then $\lambda = 1$ is an eigenvalue of multiplicity 1 and all other eigenvalues satisfy $|\lambda| < 1$