Math 252 - Linear algebra II — Nadia Lafrenière

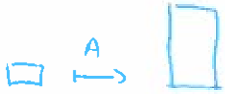
Lecture I - Eigenvalues and eigenvectors. — 16/01/2024

Goal: Decompose linear transforms by finding subspaces on which the transform acts like a scalar.

(Simple) example

Consider the transform in $\mathbb{R}^2$ that matches

$$A: (x,y) \mapsto (2x, 4y).$$



Then, $(1,0) \overset{A}{\longmapsto} (2,0) = 2\cdot(1,0)$

and $(0,1) \mapsto (0,4) = 4\cdot(0,1)$

However,

$$(1,1) \mapsto (2,4) \neq c\cdot(1,1) \text{ for any scalar } c.$$

So the horizontal line $(x,0)$ is a subspace on which A acts like a scalar (2), and so is the vertical line $(0,y)$ (with the scalar 4). No line with other directions satisfy that property.

To meet our goal, we need to find all (or a basis of them) the pairs made of a scalar (λ) and a vector ($\vec{x}$) for which the transformation A acts on $\vec{x}$ as λ: $\lambda\vec{x} = A\vec{x}$

Here λ is an eigenvalue and $\vec{x}$ is an eigenvector.

Once we found all pairs, one can decompose (diagonalize) A. (Next lecture)

Eigenvalues and eigenvectors: computation

A scalar λ (real or complex) is called an eigenvalue of the operator $A: V \rightarrow V$ if there exists a non-zero vector $\vec{x} \in V$ such that $A\vec{x} = \lambda\vec{x}$.

A vector $\vec{x}$ that satisfies $A\vec{x} = \lambda\vec{x}$ is called an eigenvector of A for the eigenvalue λ.

- The set of all eigenvectors of A for the eigenvalue λ (plus the zero vector) is called the eigenspace.

  Remark: This is a vector subspace.

- The set of all eigenvalues of A is called the spectrum, denoted $\sigma(A)$.

Computation

Since $A\vec{x} = \lambda\vec{x}$ for λ an eigenvalue and $\vec{x}$ an eigenvector, it means that

$$A\vec{x} = \lambda\vec{x} \Rightarrow A\vec{x} = (\lambda I)\vec{x} \Rightarrow A\vec{x} - \lambda I\vec{x} = \vec{0} \Rightarrow (A - \lambda I)\vec{x} = \vec{0},$$

so $\vec{x}$ is in the kernel of $A - \lambda I$.

The kernel of $A - \lambda I$ is non-trivial if and only if its determinant is 0.

Proposition

- λ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$.

  The polynomial $\det(A - \lambda I)$ is called the characteristic polynomial.

- For an eigenvalue λ of A, the eigenvectors are the solutions of $(A - \lambda I)\vec{x} = \vec{0}$.

Example

Find the characteristic polynomial, eigenvalues and eigenvectors of these two matrices.

$$A = \begin{pmatrix} 4 & -5 \\ 2 & -3 \end{pmatrix} \qquad B = \begin{pmatrix} \cos(\theta) & -\sin\theta \\ \sin(\theta) & \cos\theta \end{pmatrix}$$

rotation matrix: rotation by angle θ.

Matrix A

Characteristic polynomial:

$$\det(A-\lambda \underline{I}) = \det\begin{pmatrix} 4-\lambda & -5 \\ 2 & -3-\lambda \end{pmatrix} = (4-\lambda)(-3-\lambda)+10 = \lambda^2-\lambda-2 = (\lambda-2)(\lambda+1)$$

Eigenvalues : roots of characteristic polynomial

$$\det(A-\lambda \underline{I}) = 0 \Leftrightarrow (\lambda-2)(\lambda+1) = 0 \Leftrightarrow \lambda=-1 \text{ or } \lambda=2.$$

Eigenvectors :

$\lambda=-1$: $A-\lambda\underline{I} = \begin{pmatrix} 5 & -5 \\ 2 & -2 \end{pmatrix}$. $(A-\lambda\underline{I})\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 & -5 \\ 2 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5x_1-5x_2 \\ 2x_1-2x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$\Rightarrow x_1 = x_2$.

so $\vec{x} \in \left\{ \begin{pmatrix} x_1 \\ x_1 \end{pmatrix} \middle| \; x_1 \in \mathbb{R} \right\} = \text{span}\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}$ ← eigenspace.

$\lambda=2$ $A-\lambda\underline{I} = \begin{pmatrix} 2 & -5 \\ 2 & -5 \end{pmatrix}$ $(A-\lambda\underline{I})\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 & -5 \\ 2 & -5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1-5x_2 \\ 2x_1-5x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$\Rightarrow$ $x_2 = \frac{2}{5}x_1$, and $\vec{x} \in \text{span}\left\{\begin{pmatrix} 1 \\ 2/5 \end{pmatrix}\right\} = \text{span}\left\{\begin{pmatrix} 5 \\ 2 \end{pmatrix}\right\}$.

Matrix B

Characteristic polynomial

$$\det(B-\lambda\underline{I}) : (\cos(\theta)-\lambda)^2 + \sin^2(\theta) = \cos^2\theta + \sin^2\theta - 2\lambda\cos\theta + \lambda^2$$

$$= \lambda^2 - 2\lambda\cos\theta + 1$$

Eigenvalues

quadratic formula.

$$\frac{2\cos\theta \pm \sqrt{4\cos^2\theta - 4}}{2} = \cos\theta \pm \sqrt{\cos^2\theta - 1} = \cos\theta \pm i\sin\theta$$

Complex if $\sin(\theta) \neq 0$.

Eigenvectors.

for $\lambda = \cos\theta + i\sin\theta$

$$(B-\lambda I)\begin{pmatrix}x_1\\x_2\end{pmatrix} = \begin{pmatrix}-i\sin\theta & -\sin\theta\\ \sin\theta & -i\sin\theta\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}$$

$$= \begin{pmatrix}-ix_1-x_2\\ x_1-ix_2\end{pmatrix}\sin\theta$$ Notice that $-ix_1-x_2 = -i(x_1-ix_2)$

$$= \begin{pmatrix}0\\0\end{pmatrix} \text{ if } x_1 = ix_2.$$

so the eigenspace for $\lambda = \cos\theta + i\sin\theta$ is $\operatorname{span}\left\{\begin{pmatrix}i\\1\end{pmatrix}\right\}$.

- for $\lambda = \cos\theta - i\sin\theta$

$$(B-\lambda I)\begin{pmatrix}x_1\\x_2\end{pmatrix} = \begin{pmatrix}i\sin\theta & -\sin\theta\\ \sin\theta & i\sin\theta\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix} = \begin{pmatrix}ix_1-x_2\\ x_1+ix_2\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} \text{ if } x_2 = ix_1,$$

so the eigenspace for $\lambda = \cos\theta - i\sin\theta$ is $\operatorname{span}\left\{\begin{pmatrix}1\\i\end{pmatrix}\right\}$.

Properties

1. The number of distinct eigenvalues of a $n\times n$ matrix is at most n. Counting multiplicities, the number of eigenvalues is n.

   ↳ the highest value m such that $(x-\lambda)^m$ divides the characteristic polynomial

   Proof: the characteristic polynomial has degree n, so at most n distinct roots.

2. If $\vec{x}$ is an eigenvector and c is a non-zero scalar, then $c\vec{x}$ is an eigenvector.

   Proof: $\lambda\vec{x} = A\vec{x}$. Then $A(c\vec{x}) = c(A\vec{x}) = c\lambda\vec{x} = \lambda(c\vec{x})$.

3. Similar matrices, i.e. matrices A and B with $A = M^{-1}BM$ for a matrix M, have the same characteristic polynomial (and the same eigenvalues).

   Proof: they correspond to the same operator, in different bases.

   The eigenvalue is the "scaling factor" of the operator.

4. If the eigenvalue λ has (algebraic) multiplicity m, the dimension of the eigenspace is at most m.

↳ highest power of $(x-\lambda)$ dividing the characteristic polynomial

Remark: the dimension of the eigenspace, called the geometric multiplicity, can be less than m, as in this example.

Example

$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, characteristic polynomial is $(1-\lambda)^2$, so the spectrum is $\sigma(A) = \{1\}$. The algebraic multiplicity is 2.

Eigenspace:

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow x_2 = 0$$

So the eigenspace is $\operatorname{span}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right)$, and has dimension 1.

The geometric multiplicity is 1.

5. Let $\lambda_1, \dots, \lambda_n$ be all the eigenvalues of an $n \times n$-matrix A, counted with multiplicities. Then,

   1. $\operatorname{trace}(A) = \lambda_1 + \lambda_2 + \dots + \lambda_n$.
   2. $\det(A) = \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_n$.

Proof: exercise.

6. The eigenvalues of a triangular matrix, counting multiplicities, are the diagonal entries $a_{11}, a_{22}, \dots, a_{nn}$.

Reference: Linear Algebra Done Wrong § 4.1

Math 252 - Linear algebra II

Nadia Lafrenière

Lecture II - Diagonalization

16-18/01/2024.

Last lecture, we learned that eigenvalues and eigenvectors are useful to "decompose an operator". we learn the process of doing so when the matrix satisfies some conditions

Change of basis (reminder)

Let $b_1, b_2, \dots, b_n$ be n vectors forming a basis. Then, the change of basis to the standard basis from $\{b_1, \dots, b_n\}$ is done by the change of coordinate matrix $[I]_{\underset{\leftarrow}{SB}} = [\underset{\substack{\uparrow \\ \text{column} \\ \text{vectors}}}{b_1}, \dots, b_n]$, so that the vector $\vec{v}$ in the basis B is given by $\left([I]_{SB}\, \vec{v}\right)$ in the standard basis.

Example

Consider a basis $B = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$.

Then, $[I]_{SB} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, and $[I]_{BS} = \left([I]_{SB}\right)^{-1} = \frac{1}{3}\begin{pmatrix} -1 & 2 \\ 2 & -1 \end{pmatrix}$

The vector $(1,0)$ in the standard basis corresponds in B to

$$\frac{1}{3}\begin{pmatrix} -1 & 2 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} -1/3 \\ 2/3 \end{pmatrix}.$$

Indeed $-\frac{1}{3}\underbrace{\begin{pmatrix} 1 \\ 2 \end{pmatrix}}_{b_1} + \frac{2}{3}\underbrace{\begin{pmatrix} 2 \\ 1 \end{pmatrix}}_{b_2} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

Connection to diagonalization

If an $n \times n$-matrix A has n linearly independent eigenvectors $\{b_1, \dots, b_n\}$, we write A in that basis. Since A acts like a scalar on each of $b_1, \dots, b_n$, A is similar to a diagonal matrix (in the basis B)

Theorem

- A matrix A admits a representation $A = SDS^{-1}$, where D is a diagonal matrix and S is an invertible one if and only if there exists a basis of eigenvectors of A.
- Moreover, in this case, the diagonal entries are the eigenvalues and the columns of S are the corresponding eigenvectors.

Example

Last lecture, we saw that $A = \begin{pmatrix} 4 & -5 \\ 2 & -3 \end{pmatrix}$ has $\overset{\substack{\text{spectrum} \\ = \text{set of eigenvalues} \\ \downarrow}}{\sigma}(A) = \{-1, 2\}$

and eigenspaces $E_{-1} = \langle \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rangle$ and $E_2 = \langle \begin{pmatrix} 5 \\ 2 \end{pmatrix} \rangle$.

Then, $S = \begin{pmatrix} 1 & 5 \\ 1 & 2 \end{pmatrix}$, $D = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}$ and $S^{-1} = -\frac{1}{3} \begin{pmatrix} 2 & -5 \\ -1 & 1 \end{pmatrix}$,

so that

$$
\begin{aligned}
SDS^{-1} &= -\frac{1}{3} \begin{pmatrix} 1 & 5 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 2 & -5 \\ -1 & 1 \end{pmatrix} \\
&= -\frac{1}{3} \begin{pmatrix} 1 & 5 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} -2 & 5 \\ -2 & 2 \end{pmatrix} \\
&= -\frac{1}{3} \begin{pmatrix} -12 & 15 \\ -6 & 9 \end{pmatrix} = \begin{pmatrix} 4 & -5 \\ 2 & -3 \end{pmatrix} = A.
\end{aligned}
$$

Proof

$\Leftarrow$ If there exists a basis B of eigenvectors, we write A in that basis by writing $A = [I]_{SB} \, M \, [I]_{BS}$, for an unknown matrix M. By definition, $[I]_{SB}$ has columns that are the eigenvectors of A.

For the k-th eigenvector b_k, we know that $A b_k = \lambda_k b_k$, so the k-th row of M should be $(0, 0, \dots, \underset{\substack{\uparrow \\ k\text{-th} \\ \text{position}}}{\lambda_k}, 0, \dots, 0)$.

Hence, M is diagonal with entries $\lambda_1, \dots, \lambda_n$.

$\Rightarrow$ We need to show that if A is similar to a diagonal matrix, then it has a basis of eigenvectors.

If $A = SDS^{-1}$ with D diagonal, it means that $AS = SD$. (multiplying by S on the right).

Then, consider the coordinate vectors $e_1, \dots, e_n$:

$$AS \cdot e_k = SD \cdot e_k = S \underbrace{d_{kk}}_{\text{scalar}} e_k = d_{kk} \, S e_k .$$

Therefore, Se_k is an eigenvector of A with eigenvalue d_{kk}.

Because S is invertible, $\{Se_1, \dots, Se_n\}$ is a basis due to the fact that $\{e_1, \dots, e_n\}$ is also a basis.

▨

Application: power of a matrix.

Theorem

Let A be a diagonalizable matrix, with S be the matrix containing its eigenvectors in the columns. Then, for any integer r,

$$A^r = S D^r S^{-1} = S \begin{pmatrix} \lambda_1^r & & & 0 \\ & \lambda_2^r & & \\ & & \ddots & \\ 0 & & & \lambda_k^r \end{pmatrix} S^{-1} .$$

Proof: $A^r = (SDS^{-1})(SDS^{-1}) \dots (SDS^{-1}) = SD(S^{-1}S)D(S^{-1}S)D(S^{-1}S)D \dots DS^{-1} = SD^rS^{-1}$.

Example

Compute A^{10} for $A = \begin{pmatrix} 4 & -5 \\ 2 & -3 \end{pmatrix}$.

we already noted that $A = SDS^{-1}$, with $S = \begin{pmatrix} 1 & 5 \\ 1 & 2 \end{pmatrix}$, $D = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}$

and $S^{-1} = \frac{-1}{3} \begin{pmatrix} 2 & -5 \\ -1 & 1 \end{pmatrix}$

Then, $D^{10} = \begin{pmatrix} 1 & 0 \\ 0 & 2^{10} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1024 \end{pmatrix}$, and

④

$$A^{10} = \frac{-1}{3}\begin{pmatrix} 1 & 5 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1024 \end{pmatrix}\begin{pmatrix} 2 & -5 \\ -1 & 1 \end{pmatrix}$$

$$= \frac{-1}{3}\begin{pmatrix} 1 & 5 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 2 & -5 \\ -1024 & 1024 \end{pmatrix}$$

$$= \frac{-1}{3}\begin{pmatrix} -5118 & 5115 \\ -2046 & 2043 \end{pmatrix}$$

$$= \begin{pmatrix} 1706 & -1705 \\ 682 & -681 \end{pmatrix}.$$

<u>Reference</u> : Linear Algebra Done Wrong. § 4.2.

Math 252 - Linear Algebra II

Nadia Lafrenière

Lecture 3 - Diagonalizability

23/1/2024.

Last lecture, we saw that a matrix A admits a representation as a diagonal matrix, $A = SDS^{-1}$, if and only if there exists a basis of eigenvectors of A. We give criteria for the existence of a basis of eigenvectors, making it easy to say if a matrix is diagonalizable.

Recall that theorem:

Theorem

A matrix A admits a representation $A = SDS^{-1}$ where D is a diagonal matrix if and only if there exists a basis of eigenvectors of A.

The case of n distinct eigenvalues

Theorem

Let $\lambda_1, \dots, \lambda_r$ be distinct eignvalues of A, and let $\vec{v_1}, \dots, \vec{v_r}$ be the corresponding eigen vectors. Then, $\vec{v_1}, \dots, \vec{v_r}$ are linearly independent.

Proof

By induction on r.

Base case:

r=1: Since any non-zero vector is linearly independent and eigenvectors are non-zero vectors, $\{\vec{v}_1\}$ is a set of linearly independent vectors.

Induction hypothesis: Assume that, for $r \geq 2$, the vectors $\vec{v}_1, \dots, \vec{v}_{r-1}$ are linearly independent. we need to use the induction hypothesis in the induction step

Induction step: we need to show that $\{\vec{v}_1, \dots, \vec{v}_r\}$ are linearly independent, where these vectors are eigenvectors for the distinct eigenvalues $\lambda_1, \dots, \lambda_r$, respectively.

Recall that $\{\vec{v}_1, \dots, \vec{v}_r\}$ are linearly independent if and only if there is no $c_1, \dots, c_{r-1}, c_r$ (not all zero) with

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \dots + c_r\vec{v}_r = \vec{0}.$$

Let $c_1, \dots, c_r$ be such that $c_1\vec{v}_1 + c_2\vec{v}_2 + \dots + c_r\vec{v}_r = \vec{0}$.

Case 1: $c_r = 0$.

If $c_r = 0$, then $c_1\vec{v}_1 + \dots + c_{r-1}\vec{v}_{r-1} + \underbrace{0 \cdot \vec{v}_r}_{\vec{0}} = \vec{0}$, and using the induction step (that says that $\vec{v}_1, \dots, \vec{v}_{r-1}$ are linearly independent), this means $c_1 = \dots = c_r = 0$. Hence, $\{\vec{v}_1, \dots, \vec{v}_r\}$ are linearly independent in that case.

Case 2: $c_r \neq 0$. Then, we can solve for $\vec{v}_r$:

$$c_1\vec{v}_1 + \dots + c_{r-1}\vec{v}_{r-1} + c_r\vec{v}_r = \vec{0} \Rightarrow \vec{v}_r = \underbrace{\frac{c_1}{-c_r}}_{c_1'}\vec{v}_1 + \underbrace{\frac{c_2}{-c_r}}_{c_2'}\vec{v}_2 + \dots + \underbrace{\frac{c_{r-1}}{-c_r}}_{c_{r-1}'}\vec{v}_r$$

Since $\vec{v_r}$ is an eigenvector for the eigenvalue λ_r, then

$$(A-\lambda_r I)\vec{v_r} = \vec{0} \Rightarrow (A-\lambda_r I)(c_1'\vec{v_1} + \cdots + c_{r-1}'\vec{v_{r-1}}) = \vec{0}$$

$$\Rightarrow c_1'(A-\lambda_r I)\vec{v_1} + \cdots + c_{r-1}'(A-\lambda_r I)\vec{v_{r-1}} = \vec{0}$$

$$\Rightarrow c_1'\underbrace{(\lambda_1-\lambda_r)\vec{v_1}}_{\substack{\text{because } \vec{v_1} \text{ is an eigenvector of } A \\ \text{with eigenvalue } \lambda_1}} + \cdots + c_{r-1}'(\lambda_{r-1}-\lambda_r)\vec{v_{r-1}} = \vec{0}.$$

Because $\lambda_1, \ldots, \lambda_r$ are distinct eigenvalues, each of $\lambda_1-\lambda_r, \lambda_2-\lambda_r, \ldots, \lambda_{r-1}-\lambda_r$ are non-zero.

Because $\vec{v_1}, \ldots, \vec{v_{r-1}}$ are linearly independent,

$$c_1'(\lambda_1-\lambda_r)\vec{v_1} + \cdots + c_{r-1}'(\lambda_{r-1}-\lambda_r) = \vec{0}$$

implies that $c_1'\underbrace{(\lambda_1-\lambda_r)}_{\neq 0} = c_2'\underbrace{(\lambda_2-\lambda_r)}_{\neq 0} = \cdots = c_{r-1}'\underbrace{(\lambda_{r-1}-\lambda_r)}_{\neq 0} = 0$,

So that $c_1' = c_2' = \cdots = c_{r-1}' = 0$ and $c_1 = c_2 = \cdots = c_{r-1} = 0$ and $c_r\vec{v_r} = \vec{0} \Rightarrow c_r = 0$.
Therefore, case 2 has no solution, so we must be in case 1.
Hence, $\{\vec{v_1}, \ldots, \vec{v_r}\}$ are linearly independent.

Conclusion: For any $r \geq 1$, if $\vec{v_1}, \ldots, \vec{v_r}$ are eigenvectors for the distinct eigenvalues $\lambda_1, \ldots, \lambda_r$, then $\vec{v_1}, \ldots, \vec{v_r}$ are linearly independent.

∎

Corollary

If an nxn-matrix A has n distinct eigenvalues, A is diagonalizable.

Proof

We just showed that if A has n distinct eigenvalues, it has n linearly independent eigenvectors. Because A is an $n \times n$-matrix, it means that it lives in a space of dimension n, and therefore, the n eigenvectors form a basis.

By the theorem from last class, a matrix A is diagonalizable iff and only if there exists a basis of eigenvectors of A. Hence, A is diagonalizable.

Caveat

$$\begin{matrix} \text{n distinct eigenvalues} & \Longrightarrow & \text{diagonalizable.} \\ & \cancel{\Longleftarrow} & \end{matrix}$$

(sufficient, not necessary)

General case

we now give a necessary and sufficient criterion for diagonalizability.

Theorem

Let A be an $n \times n$-matrix. Then, A is diagonalizable if and only if, for each eigenvalue λ, the dimension of its eigenspace is equal to the (algebraic) multiplicity of λ in the characteristic polynomial of A.

Remark: The dimension of the eigenspace of λ is called its geometric multiplicity. Therefore:

$$\text{Diagonalizable} \iff \begin{matrix} \text{For each } \lambda: \\ \text{algebraic multiplicity} = \text{geometric multiplicity} \end{matrix}$$

Theorem

A diagonalizable (over the complex numbers) real matrix A admits a decomposition $A = SDS^{-1}$ where S is real and D is real and diagonal if and only if all the eigenvalues of A are real.

Example

Show that $\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$ cannot be diagonalized.

We already saw that it has a unique eigenvalue, and that its eigenspace is $\text{span}(\{\begin{pmatrix}1\\0\end{pmatrix}\})$ (of dimension 1). So the algebraic multiplicity of 1 is 2, and its geometric multiplicity is 1. Therefore, it is not diagonalizable.

Example

Show that
$$\begin{pmatrix}2 & 0 & 0 & 0\\ 0 & 2 & 0 & 0\\ 0 & 0 & 1 & 4\\ 0 & 0 & 0 & -3\end{pmatrix}$$
is diagonalizable.

We know that each eigenvalue has an eigenspace of dimension 1. Here 2, 2, 1, -3 are the eigenvalues (because the matrix is upper triangular, these are the diagonal entries) we only need to find two linearly independent eigenvectors for the eigenvalue 2. Here, $\begin{pmatrix}1\\0\\0\\0\end{pmatrix}$ and $\begin{pmatrix}0\\1\\0\\0\end{pmatrix}$ are such vectors, so the matrix is diagonalizable.

Reference: Linear Algebra Dave Wrong §4.2.

Math 252 - Linear Algebra II

Nadia Lafrenière

Lecture 4 - Inner products

23-25/1/2024

We equip vector spaces with a norm (a fancy word for distance), which will then allow us to find more canonical bases.

Definition

Let V be a vector space over the field $\overset{\mathbb{R}\text{ or }\mathbb{C}\text{, often.}}{F}$

An inner product is a function that assigns, to every pair $(\vec{x},\vec{y})$ in V a scalar in F (denoted $\langle \vec{x},\vec{y}\rangle$) satisfying the following properties:

- (Conjugate) symmetry: $\langle \vec{x},\vec{y}\rangle = \overline{\langle \vec{y},\vec{x}\rangle}$, where $\underset{\text{conjugate of } a+bi}{\overline{a+bi}} = a-bi$, and $\overline{a}=a$ for a real number.
- Linearity: $\langle a\vec{x}+b\vec{y},\vec{z}\rangle = a\langle \vec{x},\vec{z}\rangle + b\langle \vec{y},\vec{z}\rangle$ for all vectors $\vec{x},\vec{y},\vec{z}$ and all scalars a,b.
- Non-negativity: $\langle \vec{x},\vec{x}\rangle \geq 0$ for all $\vec{x}$
- Non-degeneracy: if $\langle \vec{x},\vec{x}\rangle = 0$, then $\vec{x}=\vec{0}$.

  more precisely, if and only if.

Example

We can verify that the dot product over $\mathbb{R}^n$ is an inner product. Let $\vec{x}=(x_1,\dots,x_n)$ and $\vec{y}=(y_1,\dots,y_n)$, so that $\vec{x}\cdot\vec{y}=x_1y_1+\dots+x_ny_n$.

- $\vec{y}\cdot\vec{x} = \underbrace{y_1x_1}_{=x_1y_1}+\dots+\underbrace{y_nx_n}_{=x_ny_n} = x_1y_1+\dots+x_ny_n$ (because of commutativity of real numbers, i.e. $ab=ba$ for $a,b\in\mathbb{R}$)

- $$\begin{aligned} \langle a\vec{x}+b\vec{y},\vec{z}\rangle &= (ax_1+by_1)z_1 + (ax_2+by_2)z_2 + \dots + (ax_n+by_n)z_n \\ \text{distributivity over the real numbers} \rightarrow &= ax_1z_1 + by_1z_1 + ax_2z_2 + by_2z_2 + \dots + ax_nz_n + by_nz_n \\ \text{commutativity of + over } \mathbb{R} \rightarrow &= ax_1z_1 + ax_2z_2 + \dots + ax_nz_n + by_1z_1 + by_2z_2 + \dots + by_nz_n \\ \text{distributivity} \rightarrow &= a(x_1z_1 + \dots + x_nz_n) + b(y_1z_1 + \dots + y_nz_n) \\ &= a\langle\vec{x},\vec{z}\rangle + b\langle\vec{y},\vec{z}\rangle \\ &\quad \nearrow \text{definition of } \langle\cdot,\cdot\rangle \end{aligned}$$

- $\langle\vec{x},\vec{x}\rangle = \underbrace{x_1^2}_{\geq 0} + \underbrace{x_2^2}_{\geq 0} + \dots + \underbrace{x_n^2}_{\geq 0} \geq 0$ and $\langle\vec{x},\vec{x}\rangle = 0$ if and only if $x_1=x_2=\dots=x_n=0$, so when $\vec{x}=\vec{0}$.

Example

The standard inner product over $\mathbb{C}^n$ is

$$\langle\vec{z},\vec{w}\rangle = z_1\overline{w_1} + z_2\overline{w_2} + \dots + z_n\overline{w_n} \qquad (\text{where } \overline{a+bi} = a-bi)$$

This is an inner product:

- Conjugate symmetry

$$\langle\vec{z},\vec{w}\rangle = z_1\overline{w_1} + z_2\overline{w_2} + \dots + z_n\overline{w_n}$$

$$\begin{aligned} \overline{\langle\vec{w},\vec{z}\rangle} &= \overline{w_1\overline{z_1} + w_2\overline{z_2} + \dots + w_n\overline{z_n}} \\ &= \overline{w_1\overline{z_1}} + \overline{w_2\overline{z_2}} + \dots + \overline{w_n\overline{z_n}} \quad \swarrow \quad \overline{(a+b)+i(c+d)} = \overline{a+ci} + \overline{b+di} \\ &= \overline{w_1}z_1 + \overline{w_2}z_2 + \dots + \overline{w_n}z_n \quad \longleftarrow \quad \overline{(a+bi)\overline{(c+di)}} \\ &= z_1\overline{w_1} + z_2\overline{w_2} + \dots + z_n\overline{w_n} \qquad\qquad = \overline{(a+bi)(c-di)} \\ &= \langle\vec{z},\vec{w}\rangle \qquad\qquad\qquad\qquad\qquad\quad = \overline{ac+bd+(bc-ad)i} \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad = ac+bd+(ad-bc)i \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad = (a-bi)(c+di) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad = \overline{a+bi}\,(c+di) \end{aligned}$$

The other three parts work exactly as for the dot product.

The last two products can be expressed using the Hermitian adjoint of a matrix.

Definition

A matrix A has a Hermitian adjoint $A^* = \overline{A}^T$, that is the transpose of the matrix that contains the conjugate $\overline{z}$ of each entry z.

If A is a real matrix, then $A^* = A^T$. Note that $A^{**} = A$.

Proposition

The dot product (for real vectors) and the standard inner product (for complex vectors) is

$$\langle \vec{z}, \vec{w} \rangle = \vec{w}^* \vec{z},$$

assuming $\vec{w}$ and $\vec{z}$ are column vectors.

Example

Consider the space of nxn-matrices, and define

$$\langle A, B \rangle = \operatorname{trace}(B^* A)$$

This is the Frobenius inner product, and this is an inner product.

Proof: howework.

Example

Let $f(t)$ and $g(t)$ be two elements in the space of polynomials of degree at most n. Then,

$$\langle f, g \rangle = \int_{-1}^{1} f(t)\overline{g(t)}\,dt$$

is an inner product

Partial proof:

- Conjugate symmetry

$$\langle g, f \rangle = \int_{-1}^{1} \overline{g(t)\overline{f(t)}}\,dt = \int_{-1}^{1} \overline{g(t)}\;\overline{\overline{f(t)}}\,dt$$

because of linearity of the integral

$$= \int_{-1}^{1} \overline{g(t)}\, f(t)\,dt$$

$$= \int_{-1}^{1} f(t)\,\overline{g(t)}\,dt$$

commutativity of polynomials

- Linearity follows from linearity of integrals
- Non-negativity

$$\langle f, f \rangle = \int_{-1}^{1} \underbrace{f(t)\overline{f(t)}}_{\geq 0 \text{ and } =0 \text{ iff } f(t)=0}\,dt$$

is the area under a curve above the x-axis.

for any complex $f(t)$: $(a+bi)(a-bi) = a^2+b^2$

Definition.

An inner product space is a pair made of a vector space V and an inner product on V.

The norm of an inner product space is defined by

$$\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle}.$$

Properties of inner product spaces.

i) Let $\vec{x}$ be a vector. Then $\vec{x} = \vec{0}$ if and only if $\langle \vec{x}, \vec{y} \rangle = 0$ for all $\vec{y}$.

proof of $\Leftarrow$: true for $\vec{x} = \vec{y}$ + nondegeneracy

ii) Let $\vec{x}, \vec{y}$ be vectors. Then, $\vec{x} = \vec{y}$ if and only if $\langle \vec{x}, \vec{z} \rangle = \langle \vec{y}, \vec{z} \rangle$ for all $\vec{z}$.

proof of $\Leftarrow$: Use linearity + i) with $\vec{x} - \vec{y}$

iii) If A and B are two operators $A, B : X \to Y$ such that

$$\langle A\vec{x}, \vec{y} \rangle = \langle B\vec{x}, \vec{y} \rangle \text{ for all } \vec{x} \in X, \vec{y} \in Y.$$

Then $A = B$

Proof: use ii. If $A\vec{x} = B\vec{x}$ for all $\vec{x} \in X$, then $A = B$.

iv) Cauchy-Schwarz inequality

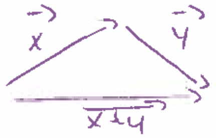
$$|\langle \vec{x}, \vec{y} \rangle| \leq \|\vec{x}\| \cdot \|\vec{y}\|$$

Proof: See textbook if interested

v) Triangle inequality

$$\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$$

Proof:



References: Linear Algebra Done Wrong, § 5.1

Math 252 - Linear Algebra II

lecture 5 - Orthogonal bases

Nadia Lafrenière

30/1/2024

Our goal for this week is to find "good" bases for some vector space. The choice of that basis depends on an inner product on the space, and should simplify computations

Orthogonality

Definition

two vectors $\vec{u}$ and $\vec{v}$ are orthogonal for the inner product $\langle \cdot , \cdot \rangle$ if $\langle \vec{u}, \vec{v} \rangle = 0$.

Notation: $\vec{u} \perp \vec{v}$.

Example: $(1,i) \perp (1,-i)$ for the standard inner product:

$$\langle (1,i), (1,-i) \rangle = 1 \cdot \bar{1} + i\,(\overline{-i})$$
$$= 1 \cdot 1 + i\,(i)$$
$$= 1 - 1 = 0.$$

Proposition

Orthogonal vectors satisfy the Pythagorean identity:

$$\|\vec{u}+\vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2,$$

where $\|\vec{u}\|$ is the norm: $\|\vec{u}\| = \sqrt{\langle \vec{u}, \vec{u} \rangle}$.

Proof

Recall that $\vec{u} \perp \vec{v}$ means $\langle \vec{u}, \vec{v} \rangle = 0$. Also, if $\langle \vec{u}, \vec{v} \rangle = 0$ means that $\langle \overline{\vec{v}}, \overline{\vec{u}} \rangle = 0$, by conjugate symmetry.

Since $\langle \overline{\vec{v}}, \overline{\vec{u}} \rangle = \overline{\langle \vec{v}, \vec{u} \rangle} = \bar{0} = 0$, then $\vec{v} \perp \vec{u}$.

Then,

$$\|\vec{u}+\vec{v}\|^2 = \langle \vec{u}+\vec{v}, \vec{u}+\vec{v} \rangle$$
$$= \langle \vec{u}, \vec{u} \rangle + \langle \vec{u}, \vec{v} \rangle + \langle \vec{v}, \vec{u} \rangle + \langle \vec{v}, \vec{v} \rangle \text{ by linearity.}$$
$$= \|\vec{u}\|^2 + 0 + 0 + \|\vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2.$$

Definition

We say that a vector $\vec{v}$ is orthogonal to a subspace W if $\vec{v} \perp \vec{w}$ for all $\vec{w} \in W$.

Lemma

Let W be spanned by $\vec{v_1}, \dots, \vec{v_r}$. Then, a vector $\vec{v}$ is orthogonal to W if and only if $\vec{v} \perp \vec{v_i}$ for all $1 \leq i \leq r$.

Proof

$\boxed{\Rightarrow}$ From the definition, if $\vec{v} \perp W$, then $\vec{v} \perp \vec{w}$ for all $\vec{w} \in W$. That includes $\vec{v_1}, \dots, \vec{v_r}$.

$\boxed{\Leftarrow}$ Let $\vec{w} \in W$. Since $\vec{v_1}, \dots, \vec{v_r}$ span W, there exist $c_1, \dots, c_r$ such that $\vec{w} = c_1 \vec{v_1} + \dots + c_r \vec{v_r}$.

Then,

$$\begin{aligned} \langle \vec{v}, \vec{w} \rangle &= \langle \vec{v}, c_1 \vec{v_1} + \dots + c_r \vec{v_r} \rangle \\ &= c_1 \cancelto{0}{\langle \vec{v}, \vec{v_1} \rangle} + \dots + c_r \cancelto{0}{\langle \vec{v}, \vec{v_r} \rangle} && \text{by linearity} \\ &= c_1 \cdot 0 + \dots + c_r \cdot 0 && \text{since } \vec{v} \perp \vec{v_i} \text{ for all } 1 \leq i \leq r \\ &= 0, \end{aligned}$$

and $\vec{v} \perp \vec{w}$.

Since $\vec{w}$ is generic within W, then $\vec{v}$ is orthogonal to all vectors in W, and $\vec{v}$ is orthogonal to W. ☒

Orthogonal systems

Definition

A system of vectors $\vec{v_1}, \dots, \vec{v_r}$ is called <u>orthogonal</u> if, for any $1 \leq i < j \leq r$, $\langle \vec{v_i}, \vec{v_j} \rangle = 0$.

An orthogonal system in which $\|\vec{v}_i\| = 1$ for all $1 \leq i \leq r$ is called orthonormal

$$\text{Orthonormal} \Rightarrow \text{orthogonal}$$
$$\not\Leftarrow$$

Lemma (Generalized Pythagorean identity)

Let $\vec{v}_1, \dots, \vec{v}_r$ be an orthogonal system.

Then,

$$\left\| \sum_{i=1}^{r} c_i \vec{v}_i \right\|^2 = \sum_{i=1}^{r} |c_i|^2 \|\vec{v}_i\|^2 .$$

Proof is analogous to that of the Pythagorean identity.

Corollary

Any orthogonal system of non-zero vectors $\vec{v}_1, \dots, \vec{v}_r$ is linearly independent.

Proof

Let $\vec{0} = c_1 \vec{v}_1 + \dots + c_r \vec{v}_r$.

We need to prove that $c_i = 0$ for all $1 \leq i \leq r$

Using the generalized Pythagorean identity,

$$0 = \|\vec{0}\|^2 = \left\| \sum_{i=1}^{r} c_i \vec{v}_i \right\|^2 = \sum_{i=1}^{r} \underbrace{|c_i|^2}_{\substack{\geq 0. \\ =0 \text{ only if} \\ c_i = 0.}} \underbrace{\|\vec{v}_i\|^2}_{\substack{\geq 0 \text{ because this is a norm.} \\ =0 \text{ only if } \vec{v}_i = \vec{0}.}} .$$

Because the $\vec{v}_i$'s are non-zero, it must be that $c_i = 0$ for all $1 \leq i \leq r$.

Hence, the vectors are linearly independent, as desired.

Definition

An orthogonal (resp. orthonormal) system that is also a basis is called an orthogonal (resp. orthonormal) basis

Remark: A system of dim(V) orthogonal $\overset{\text{non-zero}}{\vee}$ vectors in V form a basis, since they are linearly independent.

Proposition

let $\vec{v_1}, \dots, \vec{v_n}$ be an orthogonal basis for an inner product. Then, any vector $\vec{v}$ can be written in that basis as

$$\vec{v} = \sum_{k=1}^{n} \underbrace{\frac{\langle \vec{v}, \vec{v_k} \rangle}{\|\vec{v_k}\|^2}}_{\text{projection of } \vec{v} \text{ onto } \vec{v_k}} \vec{v_k}.$$

Proof

let $\vec{v} = c_1\vec{v_1} + c_2\vec{v_2} + \dots + c_n\vec{v_n}$.

then, for any $1 \le k \le n$.

$$\begin{aligned}
\langle \vec{v}, \vec{v_k} \rangle &= \langle c_1\vec{v_1} + c_2\vec{v_2} + \dots + c_n\vec{v_n}, \vec{v_k} \rangle \\
&= c_1 \overset{\to 0 \text{ if } 1 \ne k}{\langle \vec{v_1}, \vec{v_k} \rangle} + c_2 \overset{\to 0 \text{ if } 2 \ne k}{\langle \vec{v_2}, \vec{v_k} \rangle} + \dots + c_n \overset{0 \text{ if } n \ne k}{\langle \vec{v_n}, \vec{v_k} \rangle} \quad \text{by linearity} \\
&= c_1 \cdot 0 + c_2 \cdot 0 + \dots + c_{k-1} \cdot 0 + c_k \langle \vec{v_k}, \vec{v_k} \rangle + c_{k+1} \cdot 0 + \dots + 0 \quad \text{by orthogonality} \\
&= c_k \|\vec{v_k}\|^2 \quad \text{by definition of norm: } \|\vec{v_k}\| = \sqrt{\langle \vec{v_k}, \vec{v_k} \rangle}
\end{aligned}$$

by orthogonality

Therefore, $c_k = \dfrac{\langle \vec{v}, \vec{v_k} \rangle}{\|\vec{v_k}\|^2}$, for all $1 \le k \le n$

and $\vec{v} = \displaystyle\sum_{k=1}^{n} \underbrace{\frac{\langle \vec{v}, \vec{v_k} \rangle}{\|\vec{v_k}\|^2}}_{c_k} v_k$.

☒

Next class: Construction of orthogonal basis

Reference: Linear Algebra Done Wrong §5.2

Math 252 - Linear Algebra II Nadia Lafrenière

Gram-Schmidt Orthogonalization process

30/01/2024
or 01/02/2024.

We learn a process to get orthonormal and orthogonal sets of vectors spanning given subspaces.

Orthogonal projection

Definition

For a vector $\vec{v}$, its orthogonal projection onto the the subspace E, $P_E\vec{v}$, is a vector $\vec{w}$ such that

1. $\vec{w} \in E$
2. $\vec{v}-\vec{w} \perp E$.

Theorem

For any E and any $\vec{v}$,

- $P_E(P_E\vec{v}) = P_E\vec{v}$
- The projection, $P_E\vec{v}$, is unique.

Proof of uniqueness is in the textbook. We define today a procedure to do the projection.

Proposition

Let $\vec{v}_1, \ldots, \vec{v}_r$, $r\leq n$, be an orthogonal basis of $E \subseteq V$. Then,

$$P_E\vec{v} = \sum_{k=1}^{r} \frac{\langle \vec{v}, \vec{v}_k \rangle}{\|\vec{v}_k\|^2} \vec{v}_k.$$

Proof: Use proposition from last lecture, along with the fact that $\vec{v}_{r+1}, \ldots, \vec{v}_n$ are in $\underbrace{E^\perp}$, orthogonal complement of E.

Example

Project $\vec{v}=(3,2,4)$ onto $E=\text{span}(\{\underbrace{\begin{pmatrix}-1\\2\\1\end{pmatrix}}_{\vec{v}_1},\underbrace{\begin{pmatrix}2\\1\\0\end{pmatrix}}_{\vec{v}_2}\})$ (in $\mathbb{R}^3$, with dot product)

① $\left\{\begin{pmatrix}-1\\2\\1\end{pmatrix},\begin{pmatrix}2\\1\\0\end{pmatrix}\right\}$ is an orthogonal basis of E since

$$\langle(-1,2,1),(2,1,0)\rangle = -2+2+0=0.$$

② Then,

$\langle\vec{v},\vec{v}_1\rangle = \langle(3,2,4),(-1,2,1)\rangle=5$ $\quad\|\vec{v}_1\|^2=\langle\vec{v}_1,\vec{v}_1\rangle=6$

$\langle\vec{v},\vec{v}_2\rangle = \langle(3,2,4),(2,1,0)\rangle=8$ $\quad\|\vec{v}_2\|^2=\langle\vec{v}_2,\vec{v}_2\rangle=5.$

therefore,

$$P_E\vec{v}=\frac{5}{6}\begin{pmatrix}-1\\2\\1\end{pmatrix}+\frac{8}{5}\begin{pmatrix}2\\1\\0\end{pmatrix}=\frac{1}{30}\begin{pmatrix}-25+96\\50+48\\25\end{pmatrix}=\begin{pmatrix}71/30\\98/30\\25/30\end{pmatrix}$$

and

$$\vec{v}-P_E\vec{v}=\begin{pmatrix}19/30\\-38/30\\95/30\end{pmatrix}.$$

We check that $\vec{v}-P_E\vec{v}\in E^{\perp}$:

$$\left\langle\vec{v}-P_E\vec{v},\begin{pmatrix}-1\\2\\1\end{pmatrix}\right\rangle=\frac{1}{30}(-19-76+95)=0.$$

and

$$\left\langle\vec{v}-P_E\vec{v},\begin{pmatrix}2\\1\\0\end{pmatrix}\right\rangle=\frac{1}{30}(38-38)=0$$

So $P_E\vec{v}$ is indeed $\frac{1}{30}(71,98,25)$.

Example

Project (3,2,4) onto $\mathrm{span}\left(\left\{\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}0\\1\\2\end{pmatrix}\right\}\right) = E$.

Be careful! $(1,1,1) \not\perp (0,1,2)$

We first need to orthogonalize the basis for E.

① We can keep (1,1,1) as a vector.

Then, $P_{\langle(1,1,1)\rangle}(0,1,2) = \dfrac{\langle (0,1,2), (1,1,1)\rangle}{\|\langle 1,1,1\rangle\|^2}(1,1,1)$

$= \dfrac{3}{3}\cdot(1,1,1)$

and $(0,1,2) - (1,1,1) = (-1,0,1)$ is orthogonal to $(1,1,1)$.

So an orthogonal basis of that space is $\{(1,1,1), (-1,0,1)\}$

② We can apply the same process as before:

$\langle (3,2,4), (1,1,1)\rangle = 9$ $\qquad \|(1,1,1)\|^2 = 3$

$\langle (3,2,4), (-1,0,1)\rangle = 1$ $\qquad \|(-1,0,1)\|^2 = 2$.

So $P_E(3,2,4) = \dfrac{9}{3}\begin{pmatrix}1\\1\\1\end{pmatrix} + \dfrac{1}{2}\begin{pmatrix}-1\\0\\1\end{pmatrix} = \begin{pmatrix}5/2\\3\\7/2\end{pmatrix}$.

We check that $(3,2,4) - P_E(3,2,4) \in E^\perp$:

$(3,2,4) - P_E(3,2,4) = (1/2, -1, 1/2)$

$\langle (1/2, -1, 1/2), (1,1,1)\rangle = 0$ ✓

$\langle (1/2, -1, 1/2), (-1,0,1)\rangle = 0.$ ✓.

Therefore, $P_E(3,2,4) = (5/2, 3, 7/2)$.

Gram-Schmidt orthogonalization process.

Suppose we have a linearly independent system $\vec{v_1}, \dots, \vec{V_n}$.
The Gram-Schmidt method constructs an orthogonal system $\vec{w_1}, \dots, \vec{w_n}$ such that

$$\operatorname{span}(\{\vec{v_1}, \dots, \vec{v_r}\}) = \operatorname{span}(\{\vec{w_1}, \dots, \vec{w_r}\}), \text{ for all } 1 \leq r \leq n.$$

Step 1:

$\vec{w_1} = \vec{v_1}$. Define $E_1 = \operatorname{span}(\{\vec{w_1}\})$

Step 2:

$$\vec{w_2} = \underbrace{\vec{v_2} - P_{E_1}\vec{v_2}}_{\text{orthogonal to } E_1}. \quad \text{Define } E_2 = \operatorname{span}(\{\vec{w_1}, \vec{w_2}\}).$$

⋮

Step r:

$$\vec{w_r} = \underbrace{\vec{v_r} - P_{E_{r-1}}\vec{v_r}}_{\text{orthogonal to } E_{r-1}}, \quad \text{where } E_{r-1} = \operatorname{span}(\{\vec{w_1}, \dots, \vec{w_{r-1}}\})$$

Example

Construct an orthonormal basis for $\vec{v_1} = (1,1,1)$, $\vec{v_2} = (0,1,2)$, $\vec{v_3} = (1,0,2)$ using the Gram-Schmidt process

Here, $\vec{w_1} = (1,1,1)$

In the previous example, we already constructed $\vec{w_2} = (-1,0,1)$.

Then, for $\vec{w_3}$:

$$\langle \vec{w_1}, \vec{v_3} \rangle = 3, \qquad \|\vec{w_1}\|^2 = 3$$

$$\langle \vec{w_2}, \vec{v_3} \rangle = 1 \qquad \|\vec{w_2}\|^2 = 2$$

so $P_{E_2}\vec{v_3} = \frac{3}{3}\begin{pmatrix}1\\1\\1\end{pmatrix} + \frac{1}{2}\begin{pmatrix}-1\\0\\1\end{pmatrix} = \begin{pmatrix}1/2\\1\\3/2\end{pmatrix}$, and $\vec{v_3} - P_{E_2}\vec{v_3} = (1/2, -1, 1/2)$,

which we already checked in the previous example to be orthogonal to E_2.

So an orthogonal basis is

$\vec{w_1} = (1,1,1)$ $\qquad \|\vec{w_1}\| = \sqrt{3}$

$\vec{w_2} = (-1,0,1)$ $\qquad \|\vec{w_2}\| = \sqrt{2}$

$\vec{w_3} = (1/2, -1, 1/2)$ $\qquad \|\vec{w_3}\| = \sqrt{3/2}$.

An orthonormal basis is

$$\underbrace{\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)}_{\vec{w_1}/\|\vec{w_1}\|}, \underbrace{\left(\frac{-1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)}_{\vec{w_2}/\|\vec{w_2}\|}, \underbrace{\left(\frac{1}{\sqrt{6}}, \frac{-\sqrt{2}}{\sqrt{3}}, \frac{1}{\sqrt{6}}\right)}_{\vec{w_3}/\|\vec{w_3}\|}$$

Reference: Linear Algebra Done Wrong. §5.3.

Math 252 - Linear Algebra II Nadia Lafrenière

Lecture 7: Least square solution 6/2/2024

Goal: Finding the next best solution when a system $A\vec{x} = \vec{b}$ admits no solution.

Recall that the system $A\vec{x} = \vec{b}$ admits a solution only if $\vec{b}$ is in the range of A.

Example

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \vec{x} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

admits no solution, since

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + x_3 \\ x_1 + x_2 \\ x_2 - x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ a \end{pmatrix}$$

implies $a = 2 - 1 = 1$.

Therefore, the range of $\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$ is $\text{span}\left(\left\{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\right\}\right)$.

What can we do if $A\vec{x} = \vec{b}$ admits no solution?

We can minimize $A\vec{x} - \vec{b}$.

- $A\vec{x} - \vec{b} = \vec{0}$ if and only if $A\vec{x} = \vec{b}$.
- Otherwise, we want to find $\vec{x}$ such that $\|A\vec{x} - \vec{b}\|$ is minimal. This is minimal when the vector spaces $\underbrace{\text{range}(A)}_{\text{span}(A\vec{x})}$ and $\text{span}(\vec{b})$ are closest, that is, when

$$A\vec{x} = P_{\text{range}(A)}\vec{b}.$$

Theorem

The least square solution of $A\vec{x}=\vec{b}$, that is, the solution that minimizes $\|A\vec{x}-\vec{b}\|$, is given by the normal equation

$$A^*A\vec{x} = A^*\vec{b},$$

where A^* is the hermitian adjoint of A ($A^* = \bar{A}^T$)

This solution is unique if and only if A^*A is invertible

Applications: Data analysis

Say we have data coming from measurements, and we know that it should behave in some ways, but there can be errors in precision from measurements. We want to know the most likely answer, providing the solution is

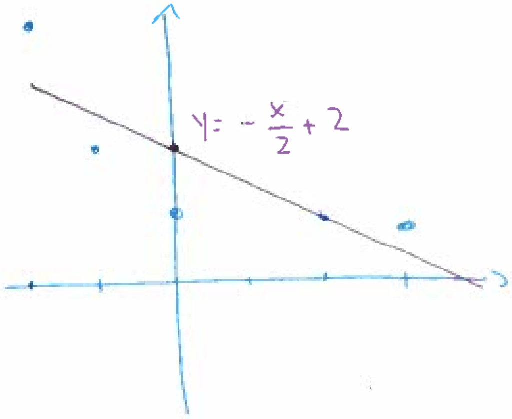
- linear : $y = ax+b$ e.g. mass vs volume
- quadratic : $y = ax^2+bx+c$ e.g. distance vs time for constant acceleration

Method: Doing least square on the model, where a, b, c are the unknown vector ($\vec{x}$), and $y_1, \cdots, y_m$ and $x_1, \cdots, x_m$ are parts of A and $\vec{b}$

Example

Do linear regression on the following data:

x	y
-2	4
-1	2
0	1
2	1
3	1



↑ if a line had the last data points, it would be horizontal

⟹ Imperfect data

To satisfy $y = ax+b$, we need to approach

$$4 = a \cdot -2 + b$$
$$2 = a \cdot -1 + b$$
$$1 = a \cdot 0 + b$$
$$1 = a \cdot 2 + b$$
$$1 = a \cdot 3 + b$$

This would mean

$$\underbrace{\begin{pmatrix} -2 & 1 \\ -1 & 1 \\ 0 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}}_{A} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

↑ $\begin{pmatrix} x_1 \\ \vdots \\ x_5 \end{pmatrix}$ ↑ all 1's

↑ $\begin{pmatrix} y_1 \\ \vdots \\ y_5 \end{pmatrix} = B$

Then,

$$A^* = \begin{pmatrix} -2 & -1 & 0 & 2 & 3 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

and

$$A^*A = \begin{pmatrix} 18 & 2 \\ 2 & 5 \end{pmatrix} \quad \text{and} \quad A^*\begin{pmatrix} y_1 \\ \vdots \\ y_5 \end{pmatrix} = \begin{pmatrix} -5 \\ 9 \end{pmatrix}.$$

Therefore,

$$\begin{pmatrix} 18 & 2 \\ 2 & 5 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 18a+2b \\ 2a+5b \end{pmatrix} = \begin{pmatrix} -5 \\ 9 \end{pmatrix},$$

and we solve this system* to get

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -1/2 \\ 2 \end{pmatrix}.$$

* by your favorite technique. For example, here, you can invert $\begin{pmatrix} 18 & 2 \\ 2 & 5 \end{pmatrix}$, and $\begin{pmatrix} 18 & 2 \\ 2 & 5 \end{pmatrix}^{-1}\begin{pmatrix} -5 \\ 9 \end{pmatrix} = \begin{pmatrix} -1/2 \\ 2 \end{pmatrix}$

Hence, the closest line to the data is

$$y = -\frac{x}{2} + 2.$$

Remark: this is what Excel does for a linear regression

Example

Suppose the same data is describing a parabola. Give the equation of the parabola that is closest to this data.

Quadratic equation: $y = ax^2 + bx + c1$ Unknowns: a, b, c

$$\underbrace{\left(\begin{array}{ccc} x^2 & x & 1 \\ 4 & -2 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{array}\right)}_{A} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \underbrace{\begin{pmatrix} 4 \\ 2 \\ 1 \\ 1 \\ 1 \end{pmatrix}}_{B}$$

Then,

$$A^*A = \begin{pmatrix} 4 & 1 & 0 & 4 & 9 \\ -2 & -1 & 0 & 2 & 3 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 4 & -2 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 114 & 26 & 18 \\ 26 & 18 & 2 \\ 18 & 2 & 5 \end{pmatrix}$$

and

$$A^*B = \begin{pmatrix} 4 & 1 & 0 & 4 & 9 \\ -2 & -1 & 0 & 2 & 3 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 2 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 31 \\ -5 \\ 9 \end{pmatrix}.$$

Therefore, we need to solve

$$\underbrace{\begin{pmatrix} 114 & 26 & 18 \\ 26 & 18 & 2 \\ 18 & 2 & 5 \end{pmatrix}}_{\text{Non-zero determinant (2464) so invertible: } \frac{1}{2464}\begin{pmatrix} 86 & -94 & -272 \\ -94 & 246 & 240 \\ -272 & 240 & 1376 \end{pmatrix}} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 31 \\ -5 \\ 9 \end{pmatrix}$$

and

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \frac{1}{2464} \begin{pmatrix} 86 & -44 & -272 \\ -94 & 246 & 240 \\ -272 & 240 & 1376 \end{pmatrix} \begin{pmatrix} 31 \\ -5 \\ 9 \end{pmatrix}$$

$$= \frac{1}{154} \begin{pmatrix} 43 \\ -124 \\ 172 \end{pmatrix}.$$

Therefore, the closest parabola is

$$y = \frac{43}{154} x^2 - \frac{124}{154} x + \frac{172}{154}.$$

Reference: Linear Algebra Done Wrong, §5.4

Math 252 - Linear Algebra II

Nadia Lafrenière

Lecture 8 - The adjoint of a linear operator

6-8/2/2024.

We revisit the hermitian adjoint of a linear operator, as well as its invariant subspaces.

Recall that the Hermitian adjoint of a matrix A is $A^* = \overline{A^T}$, and that both the standard inner product and the dot product satisfy

$$\langle \vec{u}, \vec{v} \rangle = \vec{v}^* \vec{u}.$$

Theorem (Property of adjoint matrices)

For all $\vec{u} \in \mathbb{C}^m$ and $\vec{v} \in \mathbb{C}^n$, the $m \times n$ matrix A satisfy

$$\langle A\vec{v}, \vec{u} \rangle = \langle \vec{v}, A^*\vec{u} \rangle$$

Proof

$$\langle A\vec{v}, \vec{u} \rangle = \vec{u}^* A \vec{v} = (A^*\vec{u})^* \vec{v} = \langle \vec{v}, A^*\vec{u} \rangle,$$

where the second equality is because

$$(A^*\vec{u})^* = \vec{u}^* A^{**} = \vec{u}^* A.$$

Properties : For any A, B

i. $(A+B)^* = A^* + B^*$

ii. $(cA)^* = \bar{c} A^*$, for any $c \in \mathbb{C}$

iii. $(AB)^* = B^* A^*$

iv. $A^{**} = A$

v. $\langle A\vec{v}, \vec{u} \rangle = \langle \vec{v}, A^*\vec{u} \rangle$

Fundamental subspaces

Let A be an operator acting from one inner product space V to another one, W (it may be that $V = W$)

Recall that

- the kernel, or nullspace, of A is $\{\vec{v} \in V \mid A\vec{v} = \vec{0}\} \subseteq V$
- the range, or image, of A is $\{\vec{w} \in W \mid A\vec{v} = \vec{w}$ for some $\vec{v} \in V\}$ $\subseteq W$.
- $\dim(\operatorname{Ker}(A)) + \dim(\operatorname{Range}(A)) = \dim(V)$
- $\dim(\operatorname{Range}(A)) = \dim(\operatorname{Range}(A^T))$

Theorem

i. $\operatorname{Ker}(A^*) = \operatorname{Ran}(A)^\perp$ ← The notation $W^\perp$ means the set of all vectors in a vector space orthogonal to W.

ii. $\operatorname{Ker}(A) = \operatorname{Ran}(A^*)^\perp$

iii. $\operatorname{Ker}(A^*)^\perp = \operatorname{Ran}(A)$

iv. $\operatorname{Ker}(A)^\perp = \operatorname{Ran}(A^*)$

Proof

We first prove that all four statements are equivalent, then we prove i.

i. and iii. (resp. ii and iv) are equivalent, because $W^{\perp\perp} = W$, for any $W \subseteq V$. Therefore, $\overset{\text{from i.}}{\vee}$ $\operatorname{Ker}(A^*)^\perp = \operatorname{Ran}(A)^{\perp\perp} = \operatorname{Ran}(A)$, which is iii. The others work similarly.

Also, i. and ii. are equivalent, since $A^{**} = A$, so ....

$\operatorname{Ker}(A) = \operatorname{Ker}((A^*)^*) = \operatorname{Ran}(A^*)^\perp$.

Now, we prove i.

Let $\vec{v} \in \mathrm{Ran}(A)^{\perp}$ that means that, for any $\vec{u}$,

$\langle A\vec{u}, \vec{v} \rangle = 0$.

Since $\langle A\vec{u}, \vec{v} \rangle = \langle \vec{u}, A^{*}\vec{v} \rangle$, then $\langle \vec{u}, A^{*}\vec{v} \rangle = 0$, for any $\vec{u}$, including $\vec{u} = A^{*}\vec{v}$.

Because $\langle A^{*}\vec{v}, A^{*}\vec{v} \rangle = 0$, the non-degeneracy property means that $A^{*}\vec{v} = \vec{0}$, so $\vec{v} \in \mathrm{Ker}(A^{*})$.

Conversely, if $\vec{v} \in \mathrm{Ker}(A^{*})$, $A^{*}\vec{v} = \vec{0}$ and $0 = \langle \vec{u}, A^{*}\vec{v} \rangle = \langle A\vec{u}, \vec{v} \rangle$, for all $\vec{u}$, and $\vec{v} \perp \mathrm{Ran}(A)$.

Therefore, property i is proved, and, since they are equivalent, properties ii–iv are too. ∎

Consequence

There are two orthogonal spaces for an operator $A: V \to V$:

$$V = \ker(A) \oplus \mathrm{Ran}(A^{*}),$$

and, similarly, for $A: V \to W$,

$$V \underset{\substack{\uparrow \\ \text{isomorphic}}}{\cong} \ker(A) \oplus \mathrm{Ran}(A^{*})$$

Hence, $\mathrm{Ran}(A^{*})$ describes the "essential part" of an operator.

Theorem

Let A be an $m\times n$ matrix. Then: $\ker(A) = \operatorname{Ker}(A^*A)$

Proof

This statement is of the form "set equality", so we proceed by double inclusion.

$\operatorname{Ker}(A) \subseteq \operatorname{Ker}(A^*A)$

Let $\vec{v} \in \operatorname{Ker}(A)$.

then $A\vec{v} = \vec{0}$, so $(A^*A)\vec{v} = A^*(A\vec{v}) = A^*\vec{0} = \vec{0}$, because A^* is linear. So $\vec{v} \in \operatorname{Ker}(A^*A)$.

$\operatorname{Ker}(A^*A) \subseteq \operatorname{Ker}(A)$

Let $\vec{v} \in \operatorname{Ker}(A^*A)$

Then, $A^*A\vec{v} = \vec{0}$ and

$0 = \langle A^*A\vec{v}, \vec{v} \rangle = \langle A\vec{v}, A\vec{v} \rangle$. (by property of the adjoint)

By non-degeneracy, $A\vec{v} = \vec{0}$ and $\vec{v} \in \operatorname{Ker}(A)$

Hence $\operatorname{Ker}(A) = \operatorname{Ker}(A^*A)$

$\boxtimes$

Reference: Linear Algebra Done Wrong, § 5.5

Direct sums

Definition

Let $W, W_1, \dots, W_k$ be subspaces of a vector space V such that $W_i \subseteq W$ for $1 \leq i \leq k$.

We call W the direct sum of the subspaces $W_1, \dots, W_k$, that we write $W = W_1 \oplus \dots \oplus W_k$, if

- $W = W_1 + \dots + W_k$

and

- $W_i \cap \sum_{i \neq j} W_j = \{0\}$. for each i $(1 \leq i \leq k)$.

Consequences

- For each $\vec{w} \in W$, there exists a unique linear combination $\vec{w} = c_1\vec{w_1} + \dots + c_k\vec{w_k}$, where c_i is a scalar and $w_i \in W_i$ for each $1 \leq i \leq k$.
- $\dim(W) = \sum_{i=1}^{k} \dim(W_i)$.

Examples

- If A is a diagonalizable matrix with distinct eigenvalues $\lambda_1, \dots, \lambda_k$, then $V = E_{\lambda_1} \oplus \dots \oplus E_{\lambda_k}$, where the E_{λ_i}'s are the eigenspace.
- For $A : V \to V$, $V = \ker(A) \oplus \operatorname{Ran}(A^*)$.

Math 252 - Linear Algebra II

Nadia Lafrenière

Lecture 9 - Isometries

13/2/2024

Recall that an isometry, in geometry, is a transformation that preserves angles and distances.

Examples in $\mathbb{R}^n$:

- translation
- rotation
- reflection (or symmetry)
- a composition of the above.

In linear algebra, an isometry is a linear transformation that preserves angles and distances.

⚠ a translation is not a linear transformation, since $T(O) \neq O$ if O is the origin, and T is a translation that is not the identity.

Definition

An operator $\mathcal{U}: V \to W$ is an isometry if it preserves the norm,

$$\|\mathcal{U}\vec{v}\|_W = \|\vec{v}\|_V, \quad \text{for all } \vec{v} \in V.$$

From this definition, it is clear that an isometry preserves the distances but not obvious that it preserves angles.

Theorem

An operator $\mathcal{U}: V \to W$ is an isometry if and only if it preserves the inner product:

$$\langle \mathcal{U}\vec{v}_1, \mathcal{U}\vec{v}_2 \rangle = \langle \vec{v}_1, \vec{v}_2 \rangle, \quad \text{for all } \vec{v}_1, \vec{v}_2 \in V$$

To prove it, we need the polarization lemma, that construct an inner product from the norm.

Lemma (Polarization identities)

For $\vec{u}, \vec{v} \in V$, either

- $\langle \vec{u}, \vec{v} \rangle = \frac{1}{4}\left(\|\vec{u}+\vec{v}\|^2 - \|\vec{u}-\vec{v}\|^2 \right)$ if V is a real inner product space
- $\langle \vec{u}, \vec{v} \rangle = \frac{1}{4} \sum_{\alpha=\pm 1, \pm i} \alpha \|\vec{u}+\alpha\vec{v}\|^2$ if V is a complex inner product space.

Proof

The proof is done by direct computations.

Real case:

$$\frac{1}{4}\left(\|\vec{u}+\vec{v}\|^2 - \|\vec{u}-\vec{v}\|^2 \right) = \frac{1}{4}\left(\sum_{j=1}^{n} (u_j+v_j)^2 - \sum_{j=1}^{n} (u_j-v_j)^2 \right)$$

$$= \frac{1}{4} \left(\sum_{j=1}^{n} u_j^2+v_j^2+2u_jv_j - u_j^2 - v_j^2 + 2u_jv_j \right)$$

$$= \frac{1}{\cancel{4}} \sum_{j=1}^{n} \cancel{4} u_j v_j = \langle \vec{u}, \vec{v} \rangle.$$

Complex case: Exercise.

Proof (of Theorem)

We use the polarization lemma, and we treat the real and complex cases separately.

$\boxed{\Rightarrow}$

Real case: let $\mathcal{U}$ be an isometry.

$$\langle \mathcal{U}\vec{v}_1, \mathcal{U}\vec{v}_2 \rangle \overset{\text{polarization identity.}}{=} \frac{1}{4}\left(\|\mathcal{U}\vec{v}_1 + \mathcal{U}\vec{v}_2\|^2 - \|\mathcal{U}\vec{v}_1 - \mathcal{U}\vec{v}_2\|^2 \right)$$

because $\mathcal{U}$ is linear $\rightarrow$
$$= \frac{1}{4}\left(\|\mathcal{U}(\vec{v}_1 + \vec{v}_2)\|^2 - \|\mathcal{U}(\vec{v}_1 - \vec{v}_2)\|^2 \right)$$

because $\mathcal{U}$ is an isometry $\rightarrow$
$$= \frac{1}{4}\left(\|\vec{v}_1 + \vec{v}_2\|^2 - \|\vec{v}_1 - \vec{v}_2\|^2 \right)$$

polarization identity $\longrightarrow$
$$= \langle \vec{v}_1, \vec{v}_2 \rangle$$

Complex case: let $\mathcal{U}$ be an isometry. Then,

$$\langle \mathcal{U}\vec{v}_1, \mathcal{U}\vec{v}_2 \rangle = \frac{1}{4}\Big(\|\mathcal{U}\vec{v}_1 + \mathcal{U}\vec{v}_2\|^2 - \|\mathcal{U}\vec{v}_1 - \mathcal{U}\vec{v}_2\|^2 + i\|\mathcal{U}\vec{v}_1 + i\,\mathcal{U}\vec{v}_2\|^2 - i\|\mathcal{U}\vec{v}_1 - i\,\mathcal{U}\vec{v}_2\|^2 \Big)$$

linearity + isometry $\rightarrow$
$$= \frac{1}{4}\left(\|\vec{v}_1 + \vec{v}_2\|^2 - \|\vec{v}_1 - \vec{v}_2\|^2 + i\|\vec{v}_1 + i\vec{v}_2\|^2 - i\|\vec{v}_1 - i\vec{v}_2\|^2 \right)$$
$$= \langle \vec{v}_1, \vec{v}_2 \rangle$$

$\boxed{\Leftarrow}$ If $\mathcal{U}$ preserves the inner product, then

$$\|\mathcal{U}\vec{v}\| = \sqrt{\langle \mathcal{U}\vec{v}, \mathcal{U}\vec{v} \rangle} = \sqrt{\langle \vec{v}, \vec{v} \rangle} = \|\vec{v}\|,$$

so $\mathcal{U}$ is an isometry.

The following statement allows to check easily if an operator is an isometry.

Proposition

An operator $\mathcal{U}: V \to W$ is an isometry if and only if $\mathcal{U}^*\mathcal{U} = I$.

Proof

$\boxed{\Rightarrow}$ If $\mathcal{U}$ is an isometry, then it preserves the inner product, so $\langle \mathcal{U}\vec{v}_1, \mathcal{U}\vec{v}_2 \rangle = \langle \vec{v}_1, \vec{v}_2 \rangle$ for all $\vec{v}_1, \vec{v}_2 \in V$.

Then, for all $\vec{v}_1, \vec{v}_2 \in V$,

$$\langle \vec{v}_1, \vec{v}_2 \rangle \underset{\substack{\uparrow \\ \text{isometry}}}{=} \langle \mathcal{U}\vec{v}_1, \mathcal{U}\vec{v}_2 \rangle \underset{\substack{\uparrow \\ \text{property of} \\ \text{adjoint}}}{=} \langle \mathcal{U}^*\mathcal{U}\vec{v}_1, \vec{v}_2 \rangle$$

Because $\langle \vec{v}_1, \vec{v}_2 \rangle = \langle \mathcal{U}^*\mathcal{U}\vec{v}_1, \vec{v}_2 \rangle$ for all $\vec{v}_2$, then $\vec{v}_1 = \mathcal{U}^*\mathcal{U}\vec{v}_1$ for all $\vec{v}_1 \in V$, so $\mathcal{U}^*\mathcal{U} = I$.

$\boxed{\Leftarrow}$ If $\mathcal{U}^*\mathcal{U} = Id$, then, for all $\vec{v}_1, \vec{v}_2 \in V$,

$$\langle \vec{v}_1, \vec{v}_2 \rangle = \langle \mathcal{U}^*\mathcal{U}\vec{v}_1, \vec{v}_2 \rangle = \langle \mathcal{U}\vec{v}_1, \mathcal{U}\vec{v}_2 \rangle,$$

So $\mathcal{U}$ preserves the inner product, and is therefore an isometry

☒

Definition

An isometry $\mathcal{U}: V \to W$ is called *unitary* if it is invertible.

A real unitary operator is called *orthogonal*

Proposition

An isometry is unitary if and only if $\dim(V) = \dim(W)$.

Proof sketch

- An isometry is always left invertible, since $\mathcal{U}^*\mathcal{U} = Id$
- Only square matrices are invertible, and this is when $\dim(V) = \dim(W)$.

Properties of unitary operators

Let $\mathcal{U}$ be a unitary matrix. Then

i. $\mathcal{U}^{-1} = \mathcal{U}^*$ and $\mathcal{U}^{-1}$ is also unitary

ii. If $\vec{v_1}, \dots, \vec{v_n}$ is an orthonormal basis, then $\mathcal{U}\vec{v_1}, \dots, \mathcal{U}\vec{v_n}$ is also an orthonormal basis.

iii. A product of unitary operators is unitary.

iv. The columns of $\mathcal{U}$ form an orthonormal basis

v. $|\det(\mathcal{U})| = 1$. If $\mathcal{U}$ is real, then $\det(\mathcal{U}) = \pm 1$

If complex, $\det(\mathcal{U}) = z \Rightarrow z\bar{z} = 1$.

vi. If λ is an eigenvalue of $\mathcal{U}$, then $|\lambda| = 1$

Recall that eigenvalues of real matrices can be complex, such as in the case of the rotation matrix.

Proof ideas

i. Because $\mathcal{U}^*\mathcal{U} = I$, then $\mathcal{U}^*$ is the inverse of $\mathcal{U}$. It also preserves the inner product.

ii. Because an isometry preserves the inner product:

- $\|\mathcal{U}\vec{v_1}\| = \|\vec{v_1}\| = 1$
- $\langle \mathcal{U}\vec{v_1}, \mathcal{U}\vec{v_2} \rangle = \langle \vec{v_1}, \vec{v_2} \rangle = 0$, if $\vec{v_1}, \vec{v_2}$ are orthonormal

iii. The composition of operators that preserve the inner product preserves it as well. Also, the product of invertible matrices is invertible: $(AB)^{-1} = B^{-1}A^{-1}$.

iv. Let $\vec{v_1}, \dots, \vec{v_n}$ be the columns of $\mathcal{U}$.

Because $\mathcal{U}^*\mathcal{U} = I$, then

- $\langle \vec{v_i}, \vec{v_i} \rangle = 1$ for all i

and

- $\langle \vec{v_j}, \vec{v_i} \rangle = 0$ for all $j \neq i$.

therefore, $\langle \vec{v_1}, \dots, \vec{v_n} \rangle$ is an orthonormal system that contains as many vectors as $\dim(V)$, so it is an orthonormal basis

v. Let $z = \det(\mathcal{U})$. Then, $\det(\mathcal{U}^*) = \bar{z}$ and

$$1 = \det(I) = \det(\mathcal{U}^*\mathcal{U}) = \det(\mathcal{U}^*)\det(\mathcal{U}) = \bar{z}\cdot z.$$

Therefore, $|z| = \bar{z}z = 1$.

vi. Let λ be an eigenvalue of $\mathcal{U}$. Since there exists an eigenvector $\vec{v}$ for λ, we have

$$\|\vec{v}\| = \|\mathcal{U}\vec{v}\| = \|\lambda\vec{v}\| = |\lambda|\,\|\vec{v}\|,$$

so $|\lambda| = 1$.

Note that vi. $\Rightarrow$ v., since $\det(\mathcal{U}) = \lambda_1\cdots\lambda_n$.

Examples

i. The rotation matrix in $\mathbb{R}^2$.

Let $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

Then, $$A^*A = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

so A is an isometry and is orthogonal.

Note that, because $\cos(\theta) = \cos(-\theta)$ and $\sin(\theta) = -\sin(-\theta)$, A^* is the rotation by an angle of $-\theta$. Indeed, rotating by θ and then by $-\theta$ amounts to doing nothing.

ii. Let $\vec{v}_1, \dots, \vec{v}_n$ be an orthonormal basis of V and $\vec{w}_1, \dots, \vec{w}_n$ be an orthonormal basis of W.

Then, the operator $A: V \to W$ that maps $A\vec{v}_i = \vec{w}_i$ is unitary:

- $\langle \vec{v_i}, \vec{v_j} \rangle = 0$
  and $\langle A\vec{v_i}, A\vec{v_j} \rangle = \langle \vec{w_i}, \vec{w_j} \rangle = 0$. for all $i \neq j$.
- $\langle A\vec{v_i}, A\vec{v_i} \rangle = \langle \vec{w_i}, \vec{w_i} \rangle = 1 = \langle \vec{v_i}, \vec{v_i} \rangle$ for all i.

and since any vector can be obtained as a linear combination of these, then A preserves the inner product. Also, since $\dim(V) = \dim(W) = n$, it is unitary.

Reference: Linear Algebra Done Wrong. § 5.6.

Math 252 - Linear Algebra II

Nadia Lafrenière

lecture 10 - Rigid motion in $\mathbb{R}^n$

13-15/2/2024

We are interested in geometric isometries in the Euclidean space $\mathbb{R}^n$.

Definition

A rigid motion in an inner product space V is a transformation $f: V \to V$ preserving the distance between points:

$$\|f(\vec{u}) - f(\vec{v})\| = \|\vec{u} - \vec{v}\| \quad \text{for all } \vec{u}, \vec{v} \in V.$$

Note: f does not need to be linear

Example

In $\mathbb{R}^n$, the following are rigid motion:

transformation	isometry?
translation	No
rotation	Yes, if centered at the origin.
reflection	Yes, if it fixes the origin.
composition of the above	Not if with translation.

Are there others?

Theorem

Let f be a rigid motion in $\mathbb{R}^n$, and let $T(\vec{v}) = f(\vec{v}) - f(\vec{0})$.

Then, T is an orthogonal transformation.

We will give a sketch of proof of this lemma, then we will describe all isometries in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Lemma

Let T be defined as in the theorem. Then, for all $\vec{u},\vec{v} \in \mathbb{R}^n$:

(i) $\|T(\vec{u})\| = \|\vec{u}\|$

(ii) $\|T(\vec{u})-T(\vec{v})\| = \|\vec{u}-\vec{v}\|$

(iii) $\langle T(\vec{u}), T(\vec{v})\rangle = \langle \vec{u},\vec{v}\rangle$

Proof

(i) $\|T(\vec{u})\| = \|f(\vec{u})-f(\vec{0})\| \overset{\substack{\text{definition of}\\ \text{rigid motion}}}{=} \|\vec{u}-\vec{0}\| = \|\vec{u}\|$

(ii) $\|T(\vec{u})-T(\vec{v})\| = \|f(\vec{u})-f(\vec{0})) - f(\vec{v}) + f(\vec{0})\|$

$= \|f(\vec{u})-f(\vec{v})\| = \|\vec{u}-\vec{v}\|$

(iii) We notice that in $\mathbb{R}^n$,

$$\|T(\vec{u})-T(\vec{v})\|^2 = \underbrace{\|T(\vec{u})\|^2}_{\|\vec{u}\|^2} + \underbrace{\|T(\vec{v})\|^2}_{\|\vec{v}\|^2} - 2\langle T(\vec{u}), T(\vec{v})\rangle.$$

and

$$\|\vec{u}-\vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\langle \vec{u},\vec{v}\rangle.$$

Because $\|T(\vec{u})-T(\vec{v})\| = \|\vec{u}-\vec{v}\|$ (by (ii)), and $\|T(\vec{u})\| = \|\vec{u}\|$, then $\langle T(\vec{u}), T(\vec{v})\rangle = \langle \vec{u},\vec{v}\rangle$

☒

Proof of theorem (sketch)

To prove that T is orthogonal, we need to show that

(i) T preserves distances

(ii) T is linear

(iii) T is invertible.

(i) this is statement (i) of the lemma.

(ii) We prove that $T(\vec{u}+c\vec{v}) = T(\vec{u}) + cT(\vec{v})$ by showing that $\|T(\vec{u}+c\vec{v}) - T(\vec{u}) - cT(\vec{v})\| = 0$ (straightforward but relatively long. See textbook).

(iii) We showed last time that an isometry is invertible if and only if the dimensions of the domain and codomain are equal.

Here, $T: \mathbb{R}^n \to \mathbb{R}^n$, so the domain and codomain are the same.

Also, (i) and (ii) imply that T is an isometry.

Therefore, T is orthogonal.

Isometries in $\mathbb{R}^2$ and $\mathbb{R}^3$

We now have all the tools to describe explicitly the isometries and rigid motions in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Theorem

Let T be an isometry in $\mathbb{R}^2$, and let A be its matrix in the standard basis. Then, either

(i) T is a rotation and $\det(A) = 1$

(ii) T is a reflection about a line through the origin, and $\det(A) = -1$.

Proof

We showed last time that the matrix of an isometry has orthonormal columns.

Consider the first column. Because of orthonormality, the first column is a vector $\vec{v} = (v_1, v_2)$ with $|\vec{v}| = 1$. Therefore, in $\mathbb{R}^2$, this can be described as $\vec{v} = (\cos\theta, \sin\theta)$, for some θ.

The second column being orthogonal to $\vec{v}$, there are two options: either $(\sin\theta, -\cos\theta)$ or $(-\sin\theta, \cos\theta)$.

Therefore, there are two options for A:

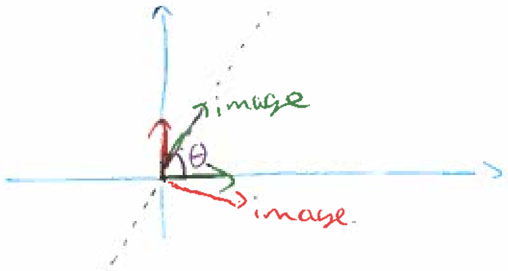
rotation matrix

$$A = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}, \text{ determinant } -1. \quad \text{or} \quad A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \text{ determinant } 1.$$

Finally, let us understand the first of these matrix.

$$A = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$$

$$A\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}, \quad A\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}\sin\theta\\ -\cos\theta\end{pmatrix}$$



This is the reflection about the axis forming an angle of $\theta/2$ with the x-axis.

Takeaway

Meaning of the theorems we saw today.

- Any rigid motion in $\mathbb{R}^n$ is the composition of a translation and an orthogonal transformation.
- The only orthogonal transformations in $\mathbb{R}^2$ are rotations and reflections.

Theorem

Let T be an orthogonal transformation in $\mathbb{R}^3$ with determinant 1. Then T is a rotation matrix.

Proof: Exercise.

Reference: Linear Algebra Done Wrong. § 5.7.

Math 252 - Linear Algebra II — Nadia Lafrenière

Lecture 11 - Upper triangular representation — 5/31/2024

After talking about when a matrix can be diagonalized, we investigate cases when they can be made triangular.

Theorem

Let $A: V \to V$ be an operator acting on a complex inner product space.

There exists an orthonormal basis $\vec{u}_1, \dots, \vec{u}_n$ such that A in that basis is upper triangular.

(In other words,

$$A = \mathcal{U} T \mathcal{U}^{-1} = \mathcal{U} T \mathcal{U}^{*}$$

where T is upper triangular, and $\mathcal{U}$ is unitary)

Proof (by induction on $n = \dim(V)$)

Base case: $n=1$. Then A is upper triangular, because all matrices are.

Induction hypothesis: For any matrix A_1 acting on a space of dimension $n-1$, there exists an orthonormal basis in which it is upper triangular.

Induction step: let A be an $n\times n$-matrix.

We need to show that A admits an upper triangular representation. We do so by dividing the matrix in four blocks:

$$A = \left(\begin{array}{c|ccc} a_{11} & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{array}\right) \begin{array}{l} \}\ 1 \\ \\ \}\ n-1 \\ \\ \end{array}$$

$$\underbrace{\quad}_{1}\ \underbrace{\qquad\qquad}_{n-1}$$

$*$ = can be anything

all need to be 0's if we want to be able to prove that A is upper triangular

$(n-1)\times(n-1)$, so induction hypothesis applies. We need A_1 to be upper triangular.

We need to show that such a division exists.

Let λ_1 be an eigenvalue of A with eigenvector $\vec{v_1}$.
Then, $A\vec{v_1} = \lambda_1 \vec{v_1}$. Let $\vec{u}_1 = \frac{v_1}{\|\vec{v_1}\|}$.

Let $E = (\operatorname{span}(\{\vec{v_1}\}))^{\perp}$. Then E is a vector space of dimension $n-1$. Let $\{\vec{v_2}, \dots, \vec{v_n}\}$ be an orthonormal basis for E. So $\{\vec{u}_1, \vec{v_2}, \vec{v_3}, \dots, \vec{v_n}\}$ is an orthonormal basis of V.

In that basis,

$$A = \left(\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{array}\right)$$

because

$$A\vec{u}_1 = \left(\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline a_{21} & & & \\ \vdots & & A_1 & \\ a_{n1} & & & \end{array}\right)\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \lambda_1 \vec{u}_1$$

because $(1,0,\dots,0)$ is an eigenvector in that basis $(\vec{u}_1)$

So $a_{21} = \dots = a_{n1} = 0$

We only need to show that A_1 can be upper triangular.

A_1 acts on E_1 which has dimension $n-1$.

By induction hypothesis, A_1 is upper triangular in some orthonormal basis $\{\vec{u_2}, \dots, \vec{u_n}\}$

Then, $\{\vec{u_1}, \dots, \vec{u_n}\}$ is an orthonormal basis in which A is upper triangular. ∎

Remark

If A is a real matrix, it may be that the triangular matrix needs to be complex

Example

: Rotation by 90°

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

Eigenvalues are i and $-i$, with eigenvectors $\begin{pmatrix} i \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ i \end{pmatrix}$ (do it in class)

Then, the only option for the triangular matrix is

$$\begin{pmatrix} i & * \\ 0 & * \end{pmatrix} \text{ or } \begin{pmatrix} -i & * \\ 0 & * \end{pmatrix}$$

The process described above gives

$$U_1 = \begin{pmatrix} 1/\sqrt{2} & i/\sqrt{2} \\ i/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \text{ for } T = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix},$$

Here, notice that the eigenvectors of $\mathcal{U}$ are orthogonal:

$$\langle (i,1),(1,i)\rangle = i\cdot\bar{1} + 1\cdot\bar{i} = i - i = 0,$$

so $\mathcal{U}$ is the matrix of eigenvectors.

Example

Consider the matrix

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

Find a unitary matrix $\mathcal{U}$ and a triangular one T such that $A = \mathcal{U} T \mathcal{U}^*$.

① Find eigenvalues and eigenvectors

(done in Homework 1):

Eigenvalues	Eigenvectors
1	(1, -1, 1)
-2	(1, 0, -1)
	(0, 1, -1)

② Set $\vec{u}_1 = \dfrac{(1,-1,1)}{\sqrt{3}}$.

Then, $(1,0,-1)$ is orthogonal to $\vec{u}_1$, so that $\vec{v}_2 = \dfrac{(1,0,-1)}{\sqrt{2}}$.

For $\vec{v}_3$, we need to find a unit vector orthogonal to both $\vec{v}_1$ and $\vec{v}_2$.

$$\vec{v_3} = \frac{(1,-1,1)\times(1,0,-1)}{\|(1,-1,1)\times(1,0,-1)\|} = \frac{(1,2,1)}{\sqrt{6}}.$$

Then, A in that basis is

$$\begin{array}{c|c} 1 & 0 \\ \hline 0 & -2 \\ 0 & 0 \end{array}$$

$\uparrow$ $A\vec{v_2}$ in basis $\vec{v_1}, \vec{v_2}, \vec{v_3}$

$\nwarrow$ $A\vec{v_3}$ in basis $\vec{v_1}, \vec{v_2}, \vec{v_3}$

$$A\vec{v_3} = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{6}}\begin{pmatrix} 10 \\ -16 \\ 10 \end{pmatrix}$$ in standard basis.

In basis $\vec{v_1}, \vec{v_2}, \vec{v_3}$, this is $c_1\vec{v_1} + c_2\vec{v_2} + c_3\vec{v_3}$ with

$$c_1 = \frac{\frac{1}{\sqrt{6}}(10,-16,10)\cdot(1,-1,1)}{\sqrt{3}} = \frac{1}{\sqrt{18}}36 = \frac{12}{\sqrt{2}}. \qquad (x\,\vec{v_1})$$

$$c_2 = \frac{1}{\sqrt{6}}\frac{(10,-16,10)\cdot(1,0,-1)}{\sqrt{2}} = 0,$$

$$c_3 = \frac{1}{\sqrt{6}}\frac{(10,-16,10)\cdot(1,2,1)}{\sqrt{6}} = -2.$$

Then A, in that basis, is

$$T = \begin{pmatrix} 1 & 0 & 12/\sqrt{2} \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix},$$

(so upper triangular).

and $$\mathcal{U} = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ -1/\sqrt{3} & 0 & 2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{pmatrix}$$

Then. $A = \mathcal{U} T \mathcal{U}^{-1} = \mathcal{U} T \mathcal{U}^*$.

Theorem

Let $A: V \to V$ be an operator acting on a real inner product space. Suppose all eigenvalues of A are real.

Then, there exists an orthogonal (real) matrix $\mathcal{U}$ and a real upper triangular matrix T such that $A = \mathcal{U} T \mathcal{U}^T$.

Proof

Analogous to the first theorem.

Reference: Linear Algebra Dave Wang §6.1.

Math 252 - Linear Algebra II

Nadia Lafrenière

Lecture 12 - Spectral Theorem

57/13/2024.

We state and prove the spectral theorem for self-adjoint matrices, and state it for normal matrices

Definition

A matrix A is self-adjoint or Hermitian if $A = A^*$.

Theorem (Spectral theorem for self-adjoint matrices)

Let $A = A^*$ be a self-adjoint matrix on a (real or complex) inner product space.

Then, all eigenvalues of A are real, and there exists an orthonormal basis of eigenvectors of A.

Also, there exists a unitary matrix $\mathcal{U}$ and a real diagonal matrix D such that

$$A = \mathcal{U} D \mathcal{U}^*.$$

Moreover, if A is a real matrix, then $\mathcal{U}$ can be chosen to be real.

We prove it after stating the following corollary:

Corollary

Let A be a real symmetric matrix.

Then, A has real eigenvalues.

Proof

Recall that there exists a unitary matrix $\mathcal{U}$ and an upper triangular matrix T such that

$$A = \mathcal{U} T \mathcal{U}^*.$$

If $A = A^*$, then

$$\mathcal{U} T \mathcal{U}^* = A = A^* = (\mathcal{U} T \mathcal{U}^*)^* = \mathcal{U} T^* \mathcal{U}^*$$

$$\Rightarrow \underbrace{\mathcal{U}^*(\mathcal{U}}_{I} T \underbrace{\mathcal{U}^*)\mathcal{U}}_{I} = \underbrace{\mathcal{U}^*(\mathcal{U}}_{I} T^* \underbrace{\mathcal{U}^*)\mathcal{U}}_{I}$$

$$\Rightarrow T = T^*.$$

We need to think about triangular matrices that are self-adjoint:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \overline{a_{12}} & a_{22} & \ddots & a_{2n} \\ \vdots & & \ddots & \vdots \\ \overline{a_{1n}} & \overline{a_{2n}} & \cdots & a_{nn} \end{pmatrix}$$

all 0's

upper triangular $\Rightarrow$ $\overline{a_{12}} = 0$ (so $a_{12} = 0$)

$\overline{a_{1n}} = 0$ ''

$\overline{a_{2n}} = 0$ '' etc.

and $a_{11} = \overline{a_{11}}$ ... $a_{nn} = \overline{a_{nn}}$

Then, a triangular matrix that is self-adjoint is diagonal, with real entries on the diagonal.

So we proved that $A = \mathcal{U} D \mathcal{U}^* = \mathcal{U} D \mathcal{U}^{-1}$.

This shows that A is similar to a diagonal matrix (so diagonalizable), and its eigenvalues are the entries of the diagonal matrix. Since D is real, then the eigenvalues of A are real.

Also, the transition matrix S such that $A = SDS^{-1}$ contains the eigenvectors of A. Therefore, the columns of $\mathcal{U}$ are eigenvectors of A, and they are an orthonormal basis.

Finally, we need to show that $\mathcal{U}$ can be chosen to be real: this follows from the last statement of the last lecture

☒

Normal matrices.

Definition

A matrix N is called normal if $N^*N = NN^*$.

Example

$$N = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$ is not self-adjoint, but it is normal since

$$NN^* = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

$$N^*N = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

So $N^*N = NN^*$, and N is normal.

Theorem (Spectral theorem for normal matrices)

Any normal operator N in a complex vector space has an orthonormal basis of eigenvectors.

In other words,

$$N = \mathcal{U} D \mathcal{U}^*,$$

where $\mathcal{U}$ is unitary and D is diagonal (D needs not to be real!)

Proof (see textbook)

Reference: Linear Algebra Dave Wrong. § 6.2.

Math 252 - Advanced Linear Algebra Nadia Lafrenière

Lectures 13-14 - Polar and Singular value decompositions

12/3/2024
(+ 14/3/2024)

Last week, we saw two versions of the spectral theorem.

We aim to expand the cases in which versions of the $\underbrace{\text{spectral theorem}}_{\star}$ apply.

Recall that, for a self-adjoint matrix A, there exists a unitary matrix $\mathcal{U}$ and a real diagonal matrix D for which

$$A = \mathcal{U} D \mathcal{U}^*$$

Also, if A is real, we can choose $\mathcal{U}$ to be real.

Understanding the meaning...

Suppose A is a $\underbrace{\text{real symmetric}}_{\text{so } A^* = A}$ matrix with nonnegative eigenvalues.

By the spectral theorem,

$$A = \mathcal{U} D \mathcal{U}^*,$$

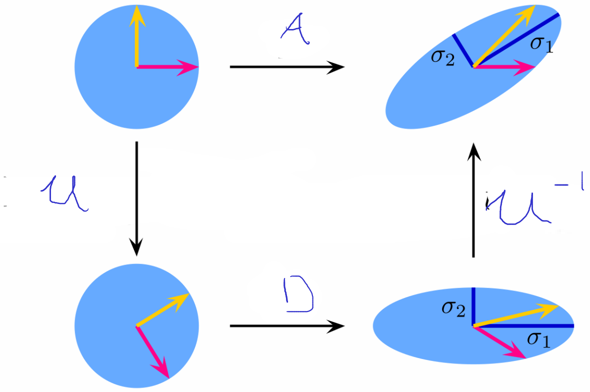
where $\mathcal{U}$ is a real isometry (rotation, reflexion, or a composition) and D is a real diagonal matrix.

The impact of D is to scale each of the coordinate vectors by a factor $\underbrace{\lambda_1}_{\geq 0}, \underbrace{\lambda_2}_{\geq 0}, \ldots, \underbrace{\lambda_n}_{\geq 0}$

by hypothesis, nonnegative eigenvalues.

In $\mathbb{R}^2$:

$\mathcal{U}$ rotates or reflects the standard basis but the result still is an orthogonal basis



Does such a decomposition of an operator exist:

- for non-symmetric matrices? Polar decomposition
- for non-square matrices? Singular value decomposition

↳ no definition of eigenvalues!

Positive (semi-) definite matrices

Definition

A self-adjoint matrix $A : V \to V$ is called positive definite if $\langle A\vec{v}, \vec{v} \rangle > 0$ for all $\vec{v} \neq \vec{0}$.

It is called positive semi-definite if $\langle A\vec{v}, \vec{v} \rangle \geq 0$ for all $\vec{v}$.

We write $A \succ 0$ for positive definite matrices, and $A \succeq 0$ for positive semi-definite.

Remark: $A > 0$ does not mean that all entries of A are positive.

⚠

Example

The matrix $\begin{pmatrix} 10 & -1 \\ -1 & 2 \end{pmatrix}$ is positive definite, since

$$\begin{pmatrix} 10 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 10v_1 - v_2 \\ -v_1 + 2v_2 \end{pmatrix},$$ for all real v1, v2

and $$\left\langle \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \begin{pmatrix} 10v_1 - v_2 \\ -v_1 + 2v_2 \end{pmatrix} \right\rangle = (10v_1^2 - v_1v_2) + (-v_1v_2 + 2v_2^2)$$

$$= 10v_1^2 + 2v_2^2 - 2v_1v_2$$

$$= 10v_1^2 + 2v_2^2 + (v_1^2 - 2v_1v_2 + v_2^2) - v_1^2 - v_2^2$$

$$= \underbrace{9v_1^2}_{\geq 0} + \underbrace{v_2^2}_{\geq 0} + \underbrace{(v_1 - v_2)^2}_{\geq 0}$$

≥ 0, and is equal to 0 only if $v_1 = v_2 = 0$.

There is a simpler way to see that a matrix is positive (semi)-definite.

Theorem

Let $A = A^*$. Then,

- $A > 0$ if and only if all eigenvalues of A are positive. (>0)
- $A \geq 0$ if and only if all eigenvalues of A are nonnegative (≥ 0)

Proof: Let $A = \mathcal{U} D \mathcal{U}^*$, with D diagonal and $\mathcal{U}$ unitary

$\boxed{\Rightarrow}$ If $A > 0$, then $\langle A\vec{v}, \vec{v} \rangle > 0$ for all $\vec{v} \neq \vec{0}$

Then, for an eigenvector $\vec{v}$,

$$\langle A\vec{v}, \vec{v} \rangle = \langle \lambda\vec{v}, \vec{v} \rangle = \lambda \underbrace{\langle \vec{v}, \vec{v} \rangle}_{>0} > 0$$

So $\lambda > 0$, and this holds for all the eigenvalues of A.

$\boxed{\Leftarrow}$ If all eigenvalues of A are positive, let us write $\vec{v} = c_1\vec{v}_1 + \dots + c_n\vec{v}_n$, where $\vec{v}_1, \dots, \vec{v}_n$ are an orthonormal basis of eigenvectors of A (it exists by the spectral theorem).

Then,

$$\langle A\vec{v}, \vec{v}\rangle = c_1\langle A\vec{v}_1, \vec{v}\rangle + c_2\langle A\vec{v}_2, \vec{v}\rangle + \dots + c_n\langle A\vec{v}_n, \vec{v}\rangle$$

$$= c_1^2 \overset{\to\lambda_1}{\cancel{\langle \lambda_1\vec{v}_1, \vec{v}_1\rangle}} + \dots + c_1c_n \overset{\to 0,\ \text{since } \langle \vec{v}_i, \vec{v}_j\rangle = 0 \text{ if } i \neq j}{\cancel{\langle \lambda_1\vec{v}_1, \vec{v}_n\rangle}} + \dots$$

$$+ c_nc_1 \overset{\to 0}{\cancel{\langle \lambda_n\vec{v}_n, \vec{v}_1\rangle}} + \dots + c_n^2 \overset{\to\lambda_n}{\cancel{\langle \lambda_n\vec{v}_n, \vec{v}_n\rangle}}$$

$$= c_1^2\lambda_1 + \dots + c_n^2\lambda_n > 0$$ (only 0 if $c_1 = \dots = c_n = 0$, so if $\vec{v} = \vec{0}$)

the proof works the same way if the eigenvalues of A are nonnegative.

∎

Corollary

Let $A = A^* \geq 0$ be a positive semidefinite operator.

There exists a unique positive semidefinite B such that $B^2 = A$. B is called the (positive) square root of A, denoted $B = \sqrt{A}$.

Proof idea

Let $A = UDU^*$, with U unitary and D a diagonal matrix with nonnegative entries. Then, choose $B = U\sqrt{D}U^*$

$$\begin{aligned} B^2 = (U\sqrt{D}U^*)^2 &= U\sqrt{D}\,\overset{I}{\cancel{U^* U}}\,\sqrt{D}U^* \\ &= U(\sqrt{D})^2 U^* \\ &= UDU^* = A. \end{aligned}$$

To find $\sqrt{D}$, recall that, for diagonal matrices,

$$\begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix}^k = \begin{pmatrix} a_{11}^k & & 0 \\ & \ddots & \\ 0 & & a_{nn}^k \end{pmatrix},$$

so $$\sqrt{D} = \begin{pmatrix} \sqrt{d_{11}} & & 0 \\ & \ddots & \\ 0 & & \sqrt{d_{nn}} \end{pmatrix}.$$

Also, since $A \geq 0$, the diagonal entries are nonnegative, and so are their square roots

∎

Modulus and singular values

⑥

Definitions

Consider an operator $A : V \to W$

Its Hermitian square is A^*A.

Observe that A^*A is a square, self-adjoint positive semidefinite matrix (proof: Homework 6)

The modulus of A is $|A| = \sqrt{A^*A}$. It is a square matrix, and it is well-defined, by the corollary above.

Proposition

For $A : V \to W$

(i) $\| A\vec{v} \| = \| \, |A| \vec{v} \, \| \quad \forall \vec{v} \in V$

(ii) $\mathrm{Ker}(A) = \ker(|A|)$.

Proof

property of the adjoint

(i) $\| \, |A| \vec{v} \, \| = \sqrt{\langle |A|\vec{v}, |A|\vec{v} \rangle} = \sqrt{\langle |A|^*|A|\vec{v}, \vec{v} \rangle}$

$= \sqrt{\langle |A|^2 \vec{v}, \vec{v} \rangle}$ because $|A|$ is self conjugate

$= \sqrt{\langle A^*A \vec{v}, \vec{v} \rangle}$ by definition of $|A|$

$= \sqrt{\langle A\vec{v}, A\vec{v} \rangle} = \| A\vec{v} \|$.

(ii) $\vec{v} \in \ker(A) \iff A\vec{v} = \vec{0} \iff \| A\vec{v} \| = 0$.
Because $\| A\vec{v} \| = \| \, |A| \vec{v} \, \|$, $\ker(A) = \ker(|A|)$.

Polar decomposition and Singular Value Decomposition.

Theorem

Let $A : V \to V$ be a square matrix

Then, A can be represented as

$$A = \mathcal{U} |A|,$$

where $\mathcal{U}$ is unitary.

The matrix $\mathcal{U}$ is unique if and only if A is invertible.

Meaning: $|A|$ is some "scaling factor".

$\mathcal{U}$ tells the "direction" of the matrix.

An example follows the definition of singular value decomposition.

Definition

Let A be an $m \times n$-matrix.

The eigenvalues of $|A|$, $\sigma_1, \dots, \sigma_n$ are the singular values of A

Remark: $\{\sigma_1^2, \dots, \sigma_n^2\} = \{\lambda_1, \dots, \lambda_n\}$, where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A^*A.

Theorem (Singular Value Decomposition)

Let A be an $m\times n$-matrix of rank r with positive singular values $\sigma_1 \geq \dots \geq \sigma_r$, and let Σ be the matrix defined by

$$\Sigma_{ij} = \begin{cases} \sigma_i & \text{if } i=j\leq r \\ 0 & \text{otherwise.} \end{cases}$$

Then, there exists a $m\times m$-unitary matrix $\mathcal{U}$ and an $n\times n$-unitary matrix V such that

$$A = \mathcal{U}\Sigma V^*.$$

This factorization is called a singular value decomposition (SVD)

An SVD is given by:

- the columns of V are orthonormal eigenvectors of A^*A.
- the first r columns of $\mathcal{U}$ are

$$\vec{u}_i := \frac{1}{\sigma_i} A\vec{v}_i.$$

the remaining columns form an orthonormal basis

Example.

Let $A = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{pmatrix}$. This matrix has rank $1 = r$.

$$A^*A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 2 & -2 \\ 2 & 2 & -2 \\ -2 & -2 & 2 \end{pmatrix}$$

Eigenvalues: $(6,0,0)$

Eigenvectors: $E_6 = \langle (1/\sqrt{3}, 1/\sqrt{3}, -1/\sqrt{3}) \rangle$

$E_0 = \langle (-1/\sqrt{6}, 2/\sqrt{6}, 1/\sqrt{6}),$

The eigenvectors of A^*A form an orthonormal basis:

$$V^* = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & -1/\sqrt{3} \\ -1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix}$$

rows are eigenvectors of A^*A

For U, we know

$$u_1 = \frac{1}{\sqrt{6}} \cdot A\vec{v}_1 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ -1/\sqrt{3} \end{pmatrix} = \frac{1}{\sqrt{18}} \begin{pmatrix} 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

u_2 is orthogonal, so we can choose $\begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$

Then,

$$U = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$$

Finally, $\Sigma = \begin{pmatrix} \sqrt{6} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

Therefore,

$$U \Sigma V^* = U \begin{pmatrix} \sqrt{2} & \sqrt{2} & -\sqrt{2} \\ 0 & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{pmatrix} = A.$$

and we found the SVD of A.

Proposition

Let $A = U\Sigma V^*$ be an SVD for a square matrix A.

then $A = \underbrace{(UV^*)}_{\text{unitary}}\underbrace{(V\Sigma V^*)}_{|A|}$ is a polar decomposition

Proof

- $(V\Sigma V^*)^2 = V\Sigma^2 V^* = \underset{\substack{\uparrow \\ \text{unitary.} \\ \text{eigenvectors of } A^*A}}{V}\,\underset{\substack{\nwarrow \text{eigenvalues} \\ \text{of } A^*A}}{D}\,V^* = A^*A.$

- Both U and V^* are unitary, so UV^* is unitary.

Example

Find a polar decomposition for $A = \begin{pmatrix} 11 & -5 \\ -2 & 10 \end{pmatrix}$.

- $A^*A = \begin{pmatrix} 11 & -2 \\ -5 & 10 \end{pmatrix}\begin{pmatrix} 11 & -5 \\ -2 & 10 \end{pmatrix} = \begin{pmatrix} 125 & -75 \\ -75 & 125 \end{pmatrix} = 25\begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix}$

  Eigenvalues: 50, 200 rank 2.

  $E_{50} = \text{span}\{(1, 1)\}$

  $E_{200} = \text{span}\{(1, -1)\}$

$V^* = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$, $\Sigma = \begin{pmatrix} 5\sqrt{2} & 0 \\ 0 & 10\sqrt{2} \end{pmatrix}$, $U = \begin{pmatrix} 3/5 & 4/5 \\ 4/5 & -3/5 \end{pmatrix}$ $\swarrow 1/5\sqrt{2}\, A\vec{v}_1$ $\nwarrow \frac{1}{10\sqrt{2}} A v_2$

So $U\Sigma V^* = \begin{pmatrix} 3 & 4 \\ 4 & -3 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 4 & -3 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix} = \begin{pmatrix} 11 & -5 \\ -2 & 10 \end{pmatrix} = A$ is a SVD

The unique polar decomposition for A is

$$A = (UV^*)(V\Sigma V^*)$$

$$= \underbrace{\begin{pmatrix} 7/5\sqrt{2} & -1/5\sqrt{2} \\ 1/5\sqrt{2} & 7/5\sqrt{2} \end{pmatrix}}_{\text{unitary}} \underbrace{\begin{pmatrix} 15/\sqrt{2} & -5/\sqrt{2} \\ -5/\sqrt{2} & 15/\sqrt{2} \end{pmatrix}}_{\text{modulus}}$$

Reference : Linear Algebra Done Wrong. § 6.3.

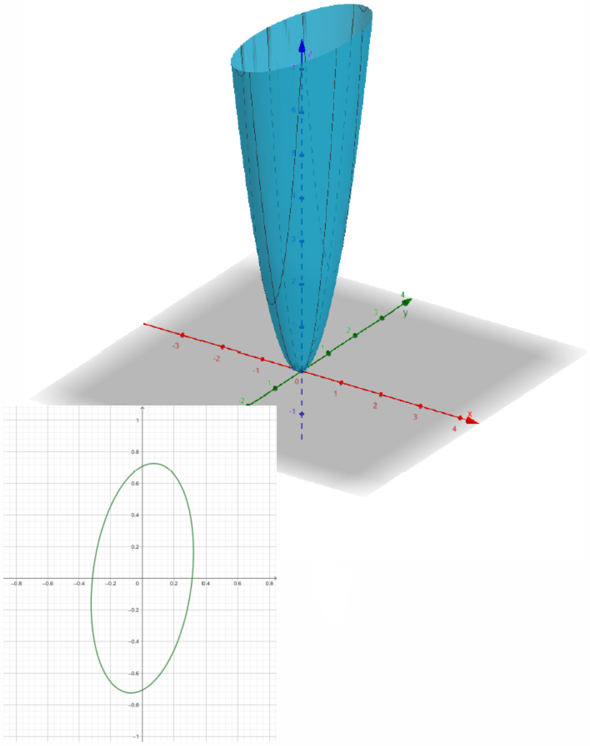
Math 252 - Advanced Linear Algebra

Nadia Lafrenière

lectures 15-16 - Quadratic forms

19-21/3/2024

We use linear algebra to understand the geometry of curves representing multivariate polynomials of degree 2.



Example

Last week, we considered the matrix

$A = \begin{pmatrix} 10 & -1 \\ -1 & 2 \end{pmatrix}$, and looked at $\langle A\vec{x}, \vec{x} \rangle$

for any real vector $\vec{x}$. We saw that

$\langle A \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} \rangle = 10x^2 + 2y^2 - 2xy$, and that

this is always positive.

Today we look at two things:

- What is the shape of $z = 10x^2 + 2y^2 - 2xy$? or $z = \langle A\vec{x}, \vec{x} \rangle$ in general?
- What curve is $c = 10x^2 + 2y^2 - 2xy$, for a fixed c?

  or $c = \langle A\vec{x}, \vec{x} \rangle$ for some c in the range of $\langle A\vec{x}, \vec{x} \rangle$.

Remark: You most likely studied conics and quadrics in a calculus course. However, the examples were usually curated so that there was no mixed terms (e.g. xy).

Examples

Conics	Equations (example)
ellipse	$\left(\frac{x}{a}\right)^2+\left(y/b\right)^2=1$
parabola	$ax^2+b=y$
hyperbola	$\left(\frac{x}{a}\right)^2-\left(\frac{y}{b}\right)^2=1$

← Not a quadratic form (term 'y' cannot appear, only xy or y^2); $y=ax^2$ is a quadratic form.

Quadrics	Equations	
Elliptic paraboloid	$ax^2+by^2=z$	$a,b>0.$
Hyperbolic paraboloid	$ax^2-by^2=z$	$a,b>0.$
Parabolic Cylinder	ax^2	$a\neq 0.$

Remarks: - For all these forms, we assume that the shape is centered at the origin.

- The main axes of the shape follow the coordinate axes.

Bilinear form

Definition

A bilinear form on $\mathbb{R}^n$ is a function $L(\vec{u},\vec{v})$, $\vec{u},\vec{v}\in\mathbb{R}^n$, such that

- $L(a\vec{u}_1+b\vec{u}_2,\vec{v}) = aL(\vec{u}_1,\vec{v})+bL(\vec{u}_2,\vec{v})$
- $L(\vec{u}, a\vec{v}_1+b\vec{v}_2) = aL(\vec{u},\vec{v}_1)+bL(\vec{u},\vec{v}_2)$

Examples

- The dot product on $\mathbb{R}^n$ is a bilinear form.
- Bilinear forms are defined similarly over $\mathbb{C}^n$. However, the standard inner product is not bilinear:

$$\text{Over } \mathbb{C}_n, \quad \langle \vec{u}, a\vec{v} \rangle = \bar{a} \langle \vec{u}, \vec{v} \rangle \neq a \langle \vec{u}, \vec{v} \rangle.$$

Over $\mathbb{R}^n$,
We can write bilinear form as matrices:

$$L(\vec{u}, \vec{v}) = \langle A\vec{u}, \vec{v} \rangle$$
$$= \sum_{i,j=1}^{n} a_{ij} u_i v_j$$

Then
$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix},$$
where A is fully determined by L.

Quadratic form

Given a bilinear form L, one defines a quadratic form as

$$Q(\vec{u}) = L(\vec{u}, \vec{u}) = \langle A\vec{u}, \vec{u} \rangle.$$

Equivalently, a quadratic form is a homogeneous polynomial of degree 2 (so the only terms are u_i^2 and $u_i u_j$).

Example Let $\vec{u} = (x, y)$ and $A = \begin{pmatrix} 10 & -1 \\ -1 & 2 \end{pmatrix}$.

Then,
$$\langle A\vec{u}, \vec{u} \rangle = \left\langle \begin{pmatrix} 10 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} \right\rangle = \left\langle \begin{pmatrix} 10x - y \\ -x + 2y \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} \right\rangle = 10x^2 - 2xy + 2y^2.$$

Example

$Q(x) = 2x^2 + 4xy - 3y^2$ is a quadratic form. Find the matrix corresponding to it.

Examples of potential answers

check them $\langle A\vec{v}, \vec{v}\rangle$

$$\begin{pmatrix} 2 & 2 \\ 2 & -3 \end{pmatrix}, \begin{pmatrix} 2 & 4 \\ 0 & -3 \end{pmatrix}, \begin{pmatrix} 2 & 10 \\ -6 & -3 \end{pmatrix}$$

Examples

The following are not quadratic forms:

- $Q(x,y) = x^2 + 2y^2 \underset{\sim}{-4}$
- $Q(x,y) = \underset{\sim}{x} - 3y^2$

Proposition

- For each quadratic form, there exist infinitely many matrix representations
- There exists a unique symmetric matrix representation:

$Q(\vec{v}) = \langle A\vec{v}, \vec{v}\rangle$ with A symmetric

Proof

Given $Q(\vec{v}) = a_{11}v_1^2 + a_{12}v_1v_2 + \dots + a_{1n}v_1v_n + a_{22}v_2^2 + \dots + a_{2n}v_2v_n + \dots + a_{nn}v_nv_n$,

we write $A = \begin{pmatrix} a_{11} & a_{12}-x_{12} \cdots & a_{1n}-x_{1n} \\ x_{12} & a_{22} & a_{2n}-x_{2n} \\ x_{1n} & x_{2n} \ddots & a_{nn} \end{pmatrix}$ for some choice of x_{ij}, $i<j$.

The only symmetric matrix is

$$\begin{pmatrix} a_{11} & a_{12}/2 & \cdots & a_{1n}/2 \\ a_{12}/2 & a_{22} & & a_{2n}/2 \\ \vdots & & \ddots & \\ a_{1n}/2 & a_{2n}/2 & & a_{nn} \end{pmatrix}$$

diagonal: a_{ii}

other entries: $a_{ij}/2$

Example

Let $f(x,y,z) = x^2 - 2xy + 6xz + z^2$

Then, the only symmetric matrix is

$$\begin{array}{c|ccc} & x & y & z \\ \hline x & 1 & -1 & 3 \\ y & -1 & 0 & 0 \\ z & 3 & 0 & 1 \end{array}$$

Reason to prefer symmetric matrices: because they are self-adjoint, they can be orthogonally diagonalized

Theorem

(Principal axis theorem)

Given a quadratic form $f(x_1, \dots, x_n)$, there exist substitutions $t_1(x_1, \dots, x_n), t_2, \dots, t_n$ such that $f(x_1, \dots, x_n) = \lambda_1 t_1^2 + \dots + \lambda_n t_n^2$.

Example

We will rewrite (soon) $f(x,y) = 2x^2 + 2y^2 + 2xy$ as

$$3/2\,(\underbrace{x+y}_{t_1})^2 + 1/2\,(\underbrace{x-y}_{t_2})^2.$$

Indeed, the latter is $\frac{1}{2}\left(3\,(x+y)^2 + (x-y)^2\right) = \frac{1}{2}\left(3x^2 + 6xy + 3y^2 + x^2 - 2xy + y^2\right)$

$= \frac{1}{2}\left(4x^2 + 4xy + 4y^2\right)$

$= 2x^2 + 2xy + 2y^2.$

Proof

let A be the matrix of the quadratic form.
Consider the matrix S such that $A = SDS^{*}$. It exists because we choose A to be symmetric such that $f(x_1, \dots, x_n) = \langle A\vec{x}, \vec{x}\rangle$, with $\vec{x} = (x_1, \dots, x_n)$. Consider $\vec{t} = S^{*} \cdot \vec{x}$ (so $\vec{x} = S\vec{t}$).

Then,

$$\langle A\vec{x}, \vec{x}\rangle = \langle AS\vec{t}, S\vec{t}\rangle = \langle S^{*}AS\vec{t}, \vec{t}\rangle$$ by property of the adjoint.

$$= \langle D\vec{t}, \vec{t}\rangle$$

$$= \lambda_1 t_1^2 + \dots + \lambda_n t_n^2 .$$

☞

Summary: $\vec{t} = S\vec{x}$ is the vector containing the appropriate substitutions

Example

Consider $f(x,y) = 2x^2 + 2xy + 2y^2$.

This is $\langle A\vec{x}, \vec{x}\rangle$ with $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.

We want to diagonalize A.

Eigenvalues:

$$\begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = \lambda^2 - 4\lambda + 3 = (\lambda-3)(\lambda-1)$$

So the eigenvalues are 1 and 3

Eigenvectors:

$\lambda = 1$: $$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y \\ x+y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \iff y = -x.$$

So $E_1 = \mathrm{Span}\left(\left\{\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right\}\right)$

$\lambda=3$ $\begin{pmatrix}-1&1\\1&-1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}-x+y\\x-y\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix} \iff x=y.$

so $E_3=\text{span}\left(\left\{\begin{pmatrix}1\\1\end{pmatrix}\right\}\right).$

Also, $(1,1)\perp(1,-1)$, so $(1/\sqrt{2}, 1/\sqrt{2})$ and $(1/\sqrt{2}, -1/\sqrt{2})$ form an orthonormal basis of eigenvectors.

Therefore, let $D=\begin{pmatrix}1&0\\0&3\end{pmatrix}$, $S=\frac{1}{\sqrt{2}}\begin{pmatrix}1&1\\-1&1\end{pmatrix}$. $\vec{t}=S^*\vec{x}=\frac{1}{\sqrt{2}}\begin{pmatrix}1&-1\\1&1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}1/\sqrt{2}(x-y)\\1/\sqrt{2}(x+y)\end{pmatrix}.$

Then,

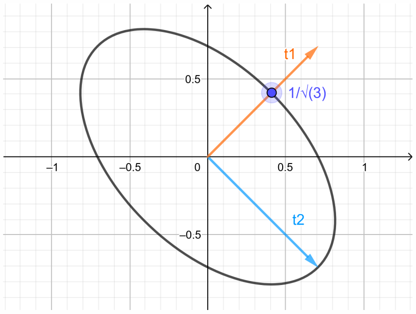
$$SDS^*=\frac{1}{2}\begin{pmatrix}1&1\\-1&1\end{pmatrix}\begin{pmatrix}1&0\\0&3\end{pmatrix}\begin{pmatrix}1&-1\\1&1\end{pmatrix}$$

$$=\frac{1}{2}\begin{pmatrix}1&1\\-1&1\end{pmatrix}\begin{pmatrix}1&-1\\3&3\end{pmatrix}=\frac{1}{2}\begin{pmatrix}4&2\\2&4\end{pmatrix}=\underbrace{\begin{pmatrix}2&1\\1&2\end{pmatrix}}_{\Rightarrow\ 2x^2+2y^2+2xy}.$$

So, $2x^2+2xy+2y^2=1\,(\underbrace{x/\sqrt{2}-\tfrac{y}{\sqrt{2}}}_{t_1})^2+3\,(\underbrace{x/\sqrt{2}+y/\sqrt{2}}_{t_2})^2$

Meaning: This quadratic form represents a rotation and scaling of x^2+3y^2. If we look at the set of points (x,y) for which $q(x,y)=1$, this is an ellipse with major axis of half-length 1 and minor axis of half length $1/\sqrt{3}$.

$$t_1^2+\left(\frac{t_2}{1/\sqrt{3}}\right)^2=1.$$



The basis is made of the unit vectors in the directions of $(1,1)$ and $(1,-1)$.

Example

Sketch $x^2 - 2\sqrt{2}\,xy = 1$.

We observe that $x^2 - 2\sqrt{2}\,xy = (x \;\; y)\begin{pmatrix} 1 & -\sqrt{2} \\ -\sqrt{2} & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$

Let $A = \begin{pmatrix} 1 & -\sqrt{2} \\ -\sqrt{2} & 0 \end{pmatrix}$ we need its eigenvalues and eigenvectors.

Eigenvalues

$$\det(A-\lambda I) = \begin{vmatrix} 1-\lambda & -\sqrt{2} \\ -\sqrt{2} & -\lambda \end{vmatrix} = -\lambda + \lambda^2 - 2 = (\lambda-2)(\lambda+1)$$

The eigenvalues are 2 and -1

Eigenvectors

$\lambda = -1$ $\begin{pmatrix} 2 & -\sqrt{2} \\ -\sqrt{2} & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x - \sqrt{2}\,y \\ -\sqrt{2}x + y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Leftrightarrow y = \sqrt{2}\,x$

So $E_{-1} = \operatorname{span}\left(\left\{\begin{pmatrix} 1 \\ \sqrt{2} \end{pmatrix}\right\}\right)$

$\lambda = 2$ $\begin{pmatrix} -1 & -\sqrt{2} \\ -\sqrt{2} & -2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -x - \sqrt{2}\,y \\ -\sqrt{2}x - 2y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Leftrightarrow x = -\sqrt{2}\,y$

so $E_2 = \operatorname{span}\left(\left\{\begin{pmatrix} -\sqrt{2} \\ 1 \end{pmatrix}\right\}\right)$

Also, these vectors are orthogonal, and they both have norm $\sqrt{3}$.

Then,

$$S = \begin{pmatrix} \sqrt{1/3} & -\sqrt{2/3} \\ \sqrt{2/3} & \sqrt{1/3} \end{pmatrix}, \quad S^* = S^T = \begin{pmatrix} \sqrt{1/3} & \sqrt{2/3} \\ -\sqrt{2/3} & \sqrt{1/3} \end{pmatrix}.$$

So $A = SDS^*$

$$= \frac{1}{3}\begin{pmatrix}1 & -\sqrt{2}\\ \sqrt{2} & 1\end{pmatrix}\begin{pmatrix}-1 & 0\\ 0 & 2\end{pmatrix}\begin{pmatrix}1 & \sqrt{2}\\ -\sqrt{2} & 1\end{pmatrix}$$

Then, it means that

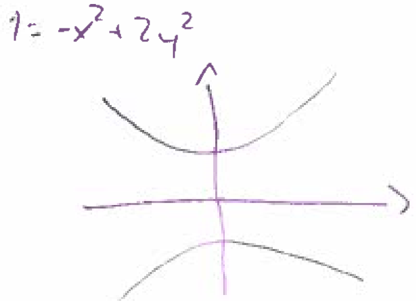
$$x^2-2\sqrt{2}xy = -\left(\frac{x+\sqrt{2}y}{\sqrt{3}}\right)^2+2\left(\frac{-\sqrt{2}x+y}{\sqrt{3}}\right)^2.$$

Indeed,

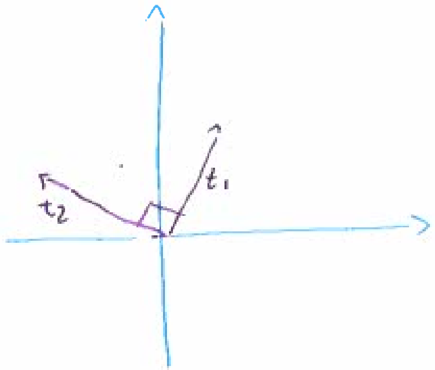
$$-\left(\frac{x+\sqrt{2}y}{\sqrt{3}}\right)^2+2\left(\frac{-\sqrt{2}x+y}{\sqrt{3}}\right)^2 = \frac{1}{3}\left(-(x^2+2\sqrt{2}xy+2y^2)+2(2x^2-2\sqrt{2}xy+y^2)\right)$$
$$=\frac{1}{3}(3x^2-6\sqrt{2}xy) = x^2-2\sqrt{2}xy.$$

To draw it, we first rewrite $t_1 = \frac{x+\sqrt{2}y}{\sqrt{3}}$ and $t_2 = \frac{-\sqrt{2}x+y}{\sqrt{3}}$

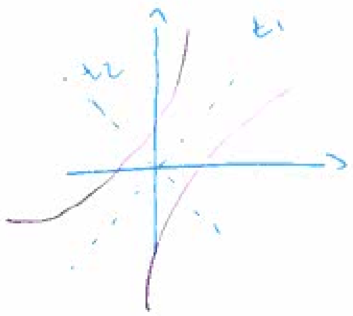
then, we have $1 = -t_1^2+2t^2$.



We also have



So the hyperbola $1 = x^2-2\sqrt{2}\,xy$ looks like



<u>Remark</u>: There is a much simpler way to recognize that this curve is a hyperbola if we don't need to know its principal axis, by completing the square to get $1 = a u_1^2 + b u_2^2$, for some u_1, u_2, with $a \cdot b < 0$

$$x^2 - 2\sqrt{2}xy = x^2 - 2\sqrt{2}xy + 2y^2 - 2y^2$$
$$= \underbrace{(x - \sqrt{2}y)}_{u_1}^2 - 2\underbrace{y}_{y=u_2}^2$$

u_1 and u_2 are not orthogonal, but we recognize $1 = u_1^2 - 2u_2$ as a hyperbola.

Reference: Linear Algebra Dave Wang §7.1 and 7.2.

Math 252 - Linear Algebra II Nadia La Frenière

Lecture 17 - Classification of quadratic forms

26/3/2024.

We learn easier ways to describe the shape of $Q(x,y,z)=1$ without having to compute principal axes.

Recall that the shapes of $Q(x,y)=1$ and $Q(x,y,z)=1$ are as follows when the principal axes are the coordinate axes:

$Q(x,y,z)=1$	Conic / quadric	
$(\frac{x}{a})^2+(\frac{y}{b})^2=1$	Ellipse	
$(\frac{x}{a})^2-(\frac{y}{b})^2=1$	Hyperbola	
$(\frac{x}{a})^2+(\frac{y}{b})^2+(\frac{z}{c})^2=1$	Ellipsoid	
$(\frac{x}{a})^2+(\frac{y}{b})^2-(\frac{z}{c})^2=1$	Hyperboloid of one sheet	
$(\frac{x}{a})^2-(\frac{y}{b})^2-(\frac{z}{c})^2=1$	Hyperboloid of two sheets	

Theorem

- Let $Q(x,y)$ be a quadratic form and let A be its symmetric matrix. Then, $Q(x,y)=1$ is the shape of
  - an ellipse, if A is positive definite
  - a hyperbola, if A is indefinite (i.e. $\langle A\vec{x},\vec{x}\rangle$ takes both positive and negative values. In other words, A has both positive and negative eigenvalues

- Let $Q(x,y,z)$ be a quadratic form and A be its symmetric matrix. Then, $Q(x,y,z)=1$ is
  - an ellipsoid if A is positive definite
  - a hyperboloid if A is indefinite with no zero eigenvalues:
    - a hyperbolaid of one sheet if it has two positive eigenvalues
    - a hyperboloid of two sheets if it has two negative eigenvalues.

Proof

This follows from the Principal axes theorem:

The theorem says that we can rewrite $Q(x_1, \dots, x_n)$ as $\lambda_1 t_1^2 + \dots + \lambda_n t_n^2$, where the λ_i's are the eigenvalues and $t_1, \dots, t_n$ correspond to orthogonal axes. Hence, the shape of the conic or the quadric is fully determined by its eigenvalues, up to a unitary transformation.

The rest of the theorem comes from comparing it with the equations on page 1.

Example

What conic is

$$x^2 - 6xy + 4xz - 6yz + 8y^2 - 3z^2 = 1 \, ?$$

We write its matrix

$$A = \begin{pmatrix} 1 & -3 & 2 \\ -3 & 8 & -3 \\ 2 & -3 & -3 \end{pmatrix}.$$

Its characteristic polynomial is $x^3 - 6x^2 - 41x + 2$.

Using, for example, a computer algebra software, we compute that its roots are roughly -4.10, 0.05 and 10.06 (they all have ugly expansions).

So this is a hyperboloid of one sheet, but we were not able to recognize it without the help of a computer. Can we do better?

Theorem (Sylvester's Law of Inertia)

Let S be any invertible matrix and D be a diagonal matrix, such that $D = S^*AS$ (Notice that S does not need to be unitary, so D and A do not need to be similar)

The number of diagonal entries that are respectively positive, negative and zero in D depend only on A, not on S.

Corollary

For a quadratic form $f(x_1, \dots, x_n)$ for which there exists $s_1(x_1, \dots, x_n), \dots, s_n(x_1, \dots, x_n)$ and numbers $d_1, \dots, d_n$ such that

$$f = d_1 s_1^2 + \dots + d_n s_n^2,$$

the number of positive (resp. negative, zero) eigenvalues of the quadratic form is the number of positive (resp. negative, zero) numbers among $d_1, \dots, d_n$.

Example

Consider again $x^2 - 6xy + 4xz - 6yz + 8y^2 - 3z^2$.

By completing the squares, we get:

$$x^2 - 6xy + 4xz - 6yz + 8y^2 - 3z^2 = (x-3y+2z)^2 - 9y^2 - 4z^2 + 12yz - 6yz + 8y^2 - 3z^2$$

$$= (x-3y+2z)^2 - y^2 + 6yz - 7z^2$$

$$= (x-3y+2z)^2 - (y-3z)^2 + 2z^2.$$

Hence, this quadratic form has two positive and one negative eigenvalues. It represents an hyperboloid of one sheet.

There even is a simpler way to know if a matrix is positive definite.

Theorem (Sylvester's Criterion for Positivity)

A matrix $A = A^*$ is positive definite if and only if

$$\det(A_k) > 0 \quad \text{for all } k = 1, \dots, n,$$

where A_k is the $k \times k$ top left submatrix of A.

Remark: This does not apply to positive semi-definite matrix:

For example, for $A = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$, $\det(A_k) = 0$ for all k, but the eigenvalues are 0 and -1.

Example

2x2 matrices.

A self-adjoint 2x2 matrix looks like

$$A = \begin{pmatrix} a & b \\ \bar{b} & c \end{pmatrix}.$$

$A_1 = (a)$ and $A_2 = A$, $ac - b\bar{b} = \det(A) = \det(A_2)$.

The eigenvalues of A, λ_1, λ_2 are both positive if and anly if both their sum and their product are positive. Their product is the determinant. Their sum is the trace:

$$\lambda_1 \lambda_2 = ac - b\bar{b}, \quad a + c = \lambda_1 + \lambda_2.$$

If $ac - b\bar{b} > 0$ and $a > 0$, c has to be positive since

$$ac > \underbrace{b\bar{b}}_{\geq 0}, \quad \text{so} \quad \underbrace{a}_{>0} c > 0.$$

Similarly, if a and c are positive, then so is $a + c$.

Proof: Linear Algebra Done Wrong. §7.3, 7.4.

Math 252 - Linear Algebra II

Nadia Lafrenière

Lecture 18 - Cayley-Hamilton Theorem & Minimal polynomial

26-28/3/2024.

Theorem (Cayley-Hamilton)

Let A be a square matrix and let $p(\lambda)$ be its characteristic polynomial (so $p(\lambda)=\det(A-\lambda I)$). Then, $p(A)=\underset{\text{matrix}}{0}$.

Example

Consider $A=\begin{pmatrix}1&1\\0&1\end{pmatrix}$. Then, $p(\lambda)=(1-\lambda)^2$, and

$$p(A)=\left(I-\begin{pmatrix}1&1\\0&1\end{pmatrix}\right)^2$$

$$=\left(\begin{pmatrix}1&0\\0&1\end{pmatrix}-\begin{pmatrix}1&1\\0&1\end{pmatrix}\right)^2$$

$$=\begin{pmatrix}0&1\\0&0\end{pmatrix}^2=\begin{pmatrix}0&0\\0&0\end{pmatrix}=0.$$

Proof (most of the idea)

We use Schur's representation theorem: $A=\mathcal{U}T\mathcal{U}^{-1}$, where T is an upper triangular matrix. Because A and T are similar, their characteristic polynomials are equal. We therefore prove the theorem for upper triangular matrices, which we can do since $p(\mathcal{U}T\mathcal{U}^{-1})=\mathcal{U}p(T)\mathcal{U}^{-1}$ for any invertible matrix $\mathcal{U}$.

For an upper triangular matrix, the eigenvalues are the diagonal entries, so

$$T=\begin{pmatrix}\lambda_1&*&\cdots&*\\&\lambda_2&\cdots&*\\&&\ddots&\vdots\\0&&&\lambda_n\end{pmatrix},$$

where $\lambda_1,\dots,\lambda_n$ are eigenvalues, and $*$ means it can be anything

Let $\vec{e_1}, \dots, \vec{e_n}$ be the vectors of the standard basis, and consider the subspaces $E_k = \text{span}(\{\vec{e_1}, \dots, \vec{e_k}\})$, of dimension k.

<u>Claim</u>: For any $\vec{v} \in E_k$, $T\vec{v} \in E_k$.

<u>Proof of claim</u>: Let $\vec{v} = c_1\vec{e_1} + \dots + c_k\vec{e_k}$. It is sufficient to prove that $T\vec{e_i} \in E_i \subseteq E_k$ for each $i \leq k$, since $T\vec{v} = c_1T\vec{e_1} + \dots + c_kT\vec{e_k}$

Because T is upper triangular,

$$T\vec{e_i} = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ & \lambda_2 & \cdots & * \\ 0 & & \ddots & \vdots \\ & & & \lambda_n \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} t_{1i} \\ t_{2i} \\ \vdots \\ t_{ii} \\ \vdots \\ t_{ni} \end{pmatrix} = \begin{pmatrix} t_{1i} \\ \vdots \\ t_{i-1,i} \\ \lambda_i \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in E_i.$$

So $T\vec{v} \in E_k$.

<u>Claim</u>: $(T - \lambda_k I)\vec{e_k} \subseteq E_{k-1}$

From the above claim, we know that $T\vec{e_k} \in E_k$. So we only need to show that the k-th coefficient of $(T - \lambda_k I)\vec{e_k}$ is 0, in other words, that the k-th coefficient of $T\vec{e_k}$ is λ_k. But this follows from above, since the entry (k,k) of T is λ_k.

To prove the theorem, it is sufficient to prove that $p(T)\vec{v} = \vec{0}$ for every vector $\vec{v}$, or even for any basis vector $\vec{e_1}, \dots, \vec{e_n}$.

Consider $\vec{e_k}$. Then

$$\underbrace{(T-\lambda_1 I)\cdots(T-\lambda_k I)}_{\text{divides } p(T)} e_k = (T-\lambda_1 I)\cdots(T-\lambda_{k-1} I)\big(\underbrace{(T-\lambda_k I)\vec{e_k}}_{\in E_{k-1}, \text{ by claim 2}}\big)$$

$$= (T-\lambda_1 I)\cdots(T-\lambda_{k-1} I)\big(\underbrace{\vec{v_{k-2}}}_{\in E_{k-2}} + a_{k-1}\vec{e_{k-1}}\big)$$

$$= (T-\lambda_1 I)\cdots(T-\lambda_{k-2} I)\big(\underbrace{(T-\lambda_{k-1} I)\vec{v_{k-2}}}_{\in E_{k-2}} + a_{k-1}\underbrace{(T-\lambda_{k-1})\vec{e_{k-1}}}_{\in E_{k-2}}\big)$$

By induction, this amounts to

$$(T-\lambda_1 I)\underbrace{\vec{v_1}}_{\in E_1} = (T-\lambda_1 I)\, a_1 \vec{e_1}$$

$$= a_1\,(T-\lambda_1 I)\,\vec{e_1} = a_1(\lambda_1 \vec{e_1} - \lambda_1 \vec{e_1}) = \vec{0}.$$

Hence, $p(T) = \underset{\text{matrix}}{\underset{\uparrow}{\mathbf{0}}}$, and $p(A) = \underset{\text{matrix}}{\underset{\uparrow}{0}}$.

Minimal polynomial

Definition

Let A be an nxn-matrix.

The minimal polynomial $q(x)$ is the polynomial of least positive degree (with leading coefficient 1) for which $q(A) = 0_{n\times n}$.

Theorem

The characteristic polynomial, $p(x)$, and the minimal polynomial, $q(x)$, have the same roots.

Proof

(i) $q(x)$ divides $p(x)$.

Consider the long division of $p(x)$ by $q(x)$. We get

$$p(x) = q(x)f(x) + \underbrace{r(x)}_{\text{degree less than } q(x)}$$

Applying it to A, this is

$$\underbrace{p(A)}_{0} = \underbrace{q(A)}_{0} f(A) + r(A),$$

so $r(x) = 0$ (because the least degree for $r(A) = 0$ is too large with $r \neq 0$).

(ii) Let λ be a root of the characteristic polynomial. In other words, λ is an eigenvalue of A, and let $\vec{v}$ be an eigenvector.

Then, $$q(A)\vec{v} = q(\lambda)\vec{v},$$

Since $A^k\vec{v} = \lambda^k\vec{v}$.

but also $q(A) = 0$, so $q(\lambda) = 0$, since $\vec{v} \neq \vec{0}$.

(iii) Let λ be a root of the minimal polynomial. Because q divides p (claim (i)), $p = q \cdot f$ and

$$p(\lambda) = \underbrace{q(\lambda)}_{0} f(\lambda) = 0,$$

so λ is also a root of p.

Corollary.

The minimal polynomial of A is

$$q(x) = (x-\lambda_1)^{m_1} \cdots (x-\lambda_k)^{m_k},$$

where $\{\lambda_1, \dots, \lambda_k\}$ are the eigenvalues of A, and m_i is an integer between 1 and the algebraic multiplicity of λ_i in A.

Example

Let $A = \begin{pmatrix} 3 & -1 & 0 \\ 0 & 2 & 0 \\ 1 & -1 & 2 \end{pmatrix}$. Compute its minimal polynomial.

Characteristic polynomial: $(3-\lambda)(2-\lambda)^2$. Eigenvalues

The minimal polynomial is either $(3-\lambda)(2-\lambda)$ or $(3-\lambda)(2-\lambda)^2$.

Attempt: $(3-\lambda)(2-\lambda) = \lambda^2 - 5\lambda + 6$

$$\Rightarrow (3-A)(2-A) = A^2 - 5A + 6$$

$$= \begin{pmatrix} 9 & -5 & 0 \\ 0 & 4 & 0 \\ 5 & -5 & 4 \end{pmatrix} - \begin{pmatrix} 15 & -5 & 0 \\ 0 & 10 & 0 \\ 5 & -5 & 10 \end{pmatrix} + \begin{pmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

So the minimal polynomial is $(3-x)(2-x)=q(x)$.

Example

Let $A=\begin{pmatrix}1&1\\0&1\end{pmatrix}$. Characteristic polynomial : $(1-\lambda)^2$.

The minimal polynomial is either $-(1-\lambda)$ or $(1-\lambda)^2$.

However, $-(I-A) = A-I = \begin{pmatrix}0&1\\0&0\end{pmatrix} \neq 0$.

So $q(x)=(1-x)^2$

Theorem

A is diagonalizable if and only if the minimal polynomial is of the form

$$q(x)=(x-\lambda_1)\cdots(x-\lambda_k),$$

where $\lambda_1,\dots,\lambda_k$ are all the distinct eigenvalues of A.

Reference: Linear Algebra Done Wrong. § 9.1

Friedberg, Insel, Spence. Linear Algebra, 5th edition. § 7.3

Math 252 - Linear Algebra II

Nadia Lafrenière

Lectures 19-20 - Jordan Form

2-4/4/2024.

We introduce the Jordan decomposition form of a matrix, which is a useful tool to describe its invariant subspaces.

Recall that a matrix is diagonalizable if and only if there exists a basis formed by its eigenvectors.

Diagonalizable <=> Existence of a basis of eigenvectors

=> Non-diagonalizable <=> "Not enough" eigenvectors.

For each eigenvalue λ there exists at least one eigenvector. The dimension of the eigenspace of λ is at most its algebraic multiplicity.

<u>Question</u>: Given an eigenvalue λ of A with algebraic multiplicity p and geometric multiplicity m, can we find $p-m$ "meaningful" vectors to complete the basis of eigenvectors?

Example

Consider $A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$. Its characteristic polynomial is $(\lambda-1)^2(\lambda-2)$. Its eigenvecters are

eigenvalue	eigenvector
2	(1,1,1)
1	(1,0,0)

From the Caylay-Hamilten theorem, we know that $(A-1)^2(A-2)\vec{v} = \vec{0}$, for all $\vec{v}$.

- We know that $(A-1)\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \vec{0}$.
- We know that $(A-2)\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \vec{0}$.
- Can we find a vector $\vec{v}$ such that $(A-1)\vec{v} \neq \vec{0}$ but $(A-1)^2\vec{v} = \vec{0}$?

  We compute $\ker((A-1)^2)$:

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} z \\ z \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Leftrightarrow z = 0.$$

  so $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \in \ker((A-1)^2) \setminus \ker(A-1)$.

We say that $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ is a generalized eigenvecter. Along with the eigenvecters, it forms a basis of the vector space with special properties.

Definition

A vector $\vec{v}$ is called a generalized eigenvector of A for eigenvalue λ if there exists m such that

$$(A-\lambda I)^m \vec{v} = \vec{0}.$$

The generalized eigenspace for eigenvalue λ is K_λ, the set of a vectors $\vec{v}$ (including $\vec{0}$) such that

$$(A-\lambda I)^m \vec{v} = \vec{0} \quad \text{for some } m.$$

Example: For $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$, the generalized eigenspaces are $K_1 = \operatorname{Span}\left(\left\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right\}\right)$ and $K_2 = \operatorname{Span}\left(\left\{\begin{pmatrix}1\\1\\1\end{pmatrix}\right\}\right)$

Theorem

Let A be a matrix with r distinct eigenvalues.

Let $b_{i1}, \dots, b_{i\dim(K_\lambda)}$ be a generalized eigenbasis of K_{λ_i}. Then,

$$\bigcup_{i=1}^{r} \{b_{i1}, \dots, b_{i,\dim(K_{\lambda_i})}\}$$ is a basis of the vector space.

union of the set

Proof: Textbook.

What is the dimension of each subspace K_{λ_i}?

Proposition

The dimension of the generalized eigenspace K_λ for an eigenvalue λ of A is its algebraic multiplicity.

Object	Eigenspace	$\subseteq$	Generalized eigenspace
Dimension	Geometric multiplicity	$\leq$	Algebraic multiplicity

Structure of the operator A.

Let K_λ be the generalized eigenspace for the eigenvalue λ of A. Then, $(A-\lambda I)\underset{\substack{\uparrow\\ \text{restriction to } K_\lambda}}{|_{K_\lambda}}$ is a nilpotent matrix, i.e. a matrix whose high powers are zero.

What does it tell about A?

Proposition

Every operator on V can be represented as $A = D + N$, where D is a diagonalizable matrix and N is a nilpotent one ($N^m = 0_{n\times n}$ for some m) such that $ND = DN$.

Proof

We decompose V into the generalized eigenspaces $K_{\lambda_1}, \dots, K_{\lambda_r}$.

On K_{λ_i}, $(A-\lambda_i I)^{m_i} = 0$, where m_i is the algebraic multiplicity of λ_i, so $N = A - \lambda_i$ is nilpotent on K_{λ_i}.

Then $A - N = \lambda_i I$, so this is the identity matrix on K_{λ_i}.

Consider a basis of generalized eigenvectors of A, and S be the matrix whose columns are the generalized eigenvectors. In that basis, one writes A as the sum of the diagonal matrix with corresponding eigenvectors, and the block matrix containing the nilpotent

$$A = S\left(\underset{D}{\overset{D'}{\begin{pmatrix} \lambda_1 \cdots \lambda_1 & & & 0 \\ & \boxed{\lambda_2 \cdots \lambda_2} & & \\ & & \ddots & \\ 0 & & & \boxed{\lambda_r \cdots \lambda_r} \end{pmatrix}}} + \overset{N'}{\begin{pmatrix} N_{\lambda_1} & & & 0 \\ & \boxed{N_{\lambda_2}} & & \\ & & \ddots & \\ 0 & & & N_{\lambda_r} \end{pmatrix}}\right) S^{-1}$$

with multiplicities; equal eigenvalues are grouped together

Then, $\underbrace{SD'S^{-1}}_{D}$ is diagonalizable and $\underbrace{SN'S^{-1}}_{N}$ is nilpotent.

We need to show that $DN=(SD'S^{-1})(SN'S^{-1})=(SN'S^{-1})(SD'S^{-1})=ND$. For this, it is sufficient to show that $D'N'=N'D'$.

On each generalized eigenspace D' is a multiple of the identity, so it commutes with any matrix: $\lambda_i I N'|_{K_{\lambda_i}} = \lambda_i N'|_{K_{\lambda_i}} I = N'|_{K_{\lambda_i}} \lambda_i I$; so $D'N'=N'D'$ and $DN=ND$.

Jordan form

Proposition

The matrix N is nilpotent if and only if there exists a basis in which it can be expressed as

$$\begin{pmatrix} 0 & * & 0 & 0 & \cdots & 0 \\ 0 & 0 & * & 0 & \cdots & 0 \\ 0 & 0 & 0 & * & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

where all the entries are 0, except maybe the ones right above the diagonal.

The proof is omitted, but we can verify that

$$n\left\{\underbrace{\begin{pmatrix}0&1&0&\cdots&0\\0&0&1&\cdots&0\\\vdots&\vdots&&\ddots&\vdots\\0&0&0&\cdots&1\\0&0&0&\cdots&0\end{pmatrix}}_{n}\right.$$

is nilpotent of order n, since

$$\begin{pmatrix}0&1&0&\cdots&0\\0&0&1&\cdots&0\\\vdots&&&&\\0&0&0&\cdots&1\\0&0&0&\cdots&0\end{pmatrix}^2=\begin{pmatrix}0&0&1&\cdots&0\\0&0&0&\ddots&0\\&&&&1\\0&0&0&\cdots&0\\0&0&0&\cdots&0\end{pmatrix},$$

and each iteration moves the non-zero diagonal by one position, until it leaves the matrix

Theorem (Jordan Decomposition Theorem)

Given a matrix A there exists a basis (the Jordan canonical basis) such that the matrix of A in this basis is

$$\begin{pmatrix}\boxed{M_{\lambda_1}}&&&\\&\boxed{M_{\lambda_2}}&&\\&&\boxed{\ddots}&\\&&&\boxed{\ }\end{pmatrix}M_{\lambda_r}\ ,$$

where M_{λ_i} is a block diagonal matrix of size between 1 and the algebraic multiplicity of λ_i, and the number of blocks for the eigenvalue λ is its geometric multiplicity:

$$M_{\lambda_i}=\begin{pmatrix}\lambda_i&1&0&\cdots&0\\0&\lambda_i&1&\cdots&0\\\vdots&\vdots&&\ddots&\vdots\\0&0&0&\cdots&\lambda_i\end{pmatrix}.$$

(a block of size 1×1 is just λ_i)

The vectors in the Jordan canonical basis are the generalized eigenvectors, where the vectors in a given generalized eigenspace are ordered in increasing value for the minimal m such that $(A-\lambda I)^m \vec{v} = \vec{0}$, and are selected such that they correspond to $\{(A-\lambda I)^m \vec{v},\ 0 \leq m <$ largest power of $x-\lambda$ in the minimal polynomial$\}$, for a fixed $\vec{v}$ that is a generalized eigenvector with highest m such that $(A-\lambda I)^m \vec{v} = \vec{0}$

Example

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$ we computed the eigenvalues and eigenvectors:

eigenvalues	eigenvectors	generalized eigenvectors
2 (mult. 1)	(1,1,1)	same
1 (mult. 2)	(1,0,0)	also (0,1,0), and $(A-I)\begin{pmatrix}0\\1\\0\end{pmatrix} = \begin{pmatrix}1\\0\\0\end{pmatrix}$

$$A = S \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} S^{-1}, \quad \text{with } S = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

Example

Let $A = \begin{pmatrix} 5 & 4 & 2 & 1 \\ 0 & 1 & -1 & -1 \\ -1 & -1 & 3 & 0 \\ 1 & 1 & -1 & 2 \end{pmatrix}$ Its charalteristic polynomial is $x^4 - 11x^3 + 42x^2 - 64x + 32$, and its eigenvectors are all multiples of $(1,-1,0,1)$, $(1,-1,0,0)$ and $(1,0,-1,1)$

1. Find its eigenvalues, with multiplicity.

2. Find a generalized eigenvector for the eigenvalue with multiplicity 2.

3. Find its Jordan decomposition.

1. Eigenvalues are obtained from eigenvectors.

$$A\begin{pmatrix}1\\-1\\0\\1\end{pmatrix}=2\begin{pmatrix}1\\-1\\0\\1\end{pmatrix},\quad A\begin{pmatrix}1\\-1\\0\\0\end{pmatrix}=\begin{pmatrix}1\\-1\\0\\0\end{pmatrix},\quad A\begin{pmatrix}1\\0\\-1\\1\end{pmatrix}=4\begin{pmatrix}1\\0\\-1\\1\end{pmatrix}.$$

The eigenvalues are 1, 2, 4.

The product of the eigenvalues are 32 (from the characteristic polynomial), so 4 is the eigenvalue with multiplicity 2.

2. $$(A-4I)^2\begin{pmatrix}w\\x\\y\\z\end{pmatrix}=\begin{pmatrix}1&4&2&1\\0&-3&-1&-1\\-1&-1&-1&0\\1&1&-1&-2\end{pmatrix}^2\begin{pmatrix}w\\x\\y\\z\end{pmatrix}=\begin{pmatrix}0&-9&-5&-5\\0&9&5&5\\0&0&0&0\\0&0&4&4\end{pmatrix}\begin{pmatrix}w\\x\\y\\z\end{pmatrix}$$

$$=\begin{pmatrix}-9x-5y-5z\\9x+5y+5z\\0\\4y+4z\end{pmatrix}=\begin{pmatrix}0\\0\\0\\0\end{pmatrix}\iff \begin{matrix}y=-z\\x=0\end{matrix}$$

so $K_4=\operatorname{span}\left(\left\{\begin{pmatrix}1\\0\\0\\0\end{pmatrix},\underset{\nwarrow\text{eigenvector}}{\begin{pmatrix}1\\0\\-1\\1\end{pmatrix}}\right\}\right)$ $\qquad (A-4I)\begin{pmatrix}1\\0\\0\\0\end{pmatrix}=\begin{pmatrix}1\\0\\-1\\1\end{pmatrix}$

So the Jordan decomposition is

$$A = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & -1 & -1 \\ -1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}}_{S} \begin{pmatrix} 4 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} S^{-1}.$$

Reference: Linear Algebra Done Wrong. § 9.3 to 9.5.

Wikipedia. Jordan normal form.

Extra example Jordan Decomposition.

let $A = \begin{pmatrix} 0 & -1 & -1 \\ -3 & -1 & -2 \\ 7 & 5 & 6 \end{pmatrix}$.

Eigenvalues	Eigenvectors
1 (algebraic multiplicity 1)	$(0,1,-1)$ (geometric multiplicity 1)
2 (algebraic multiplicity 2)	$(1,1,-3)$. (geometric multiplicity 1)

We already know J from this observation:

$$J = \left(\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & 2 & 1 \\ 0 & 0 & 2 \end{array}\right),$$

where we notice that there is a unique block for the eigenvalue 2 because the geometric multiplicity gives the number of blocks.

We need to find one generalized eigenvector for the eigenvalue 2.

$$(A-2I)^2\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}0&0&0\\1&2&1\\-1&-2&-1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}0\\x+2y+z\\-x-2y-z\end{pmatrix}=\vec{0}$$

$$\Longleftrightarrow x+2y+z=0.$$

So K_2 is the 2-dimensional space spanned by

$\begin{pmatrix}1\\0\\-1\end{pmatrix}$ and $\begin{pmatrix}0\\1\\-2\end{pmatrix}$.

For the matrix S, the third column is a vector of K_2 that is not an eigenvector. For example, choose $\begin{pmatrix}1\\0\\-1\end{pmatrix}$. Then

$$S=\begin{pmatrix}0&?&1\\1&?&0\\-1&?&-1\end{pmatrix}$$

$\uparrow$ eigenvector for 1

To find the second column, this is the eigenvector for 2 (so a multiple of $(1,1,-3)$) that is obtained from $(A-2I)\begin{pmatrix}1\\0\\-1\end{pmatrix}=\begin{pmatrix}-2&-1&-1\\-3&-3&-2\\7&5&4\end{pmatrix}\begin{pmatrix}1\\0\\-1\end{pmatrix}=\begin{pmatrix}-1\\-1\\3\end{pmatrix}$.

Therefore, the Jordan decomposition of A is $A = SJS^{-1}$, with

$$S = \begin{pmatrix} 0 & -1 & 1 \\ 1 & -1 & 0 \\ -1 & 3 & -1 \end{pmatrix} \quad \text{and} \quad J = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$