

MATH240 Notes

Yash Thakur

Contents

1	Section 1	2
1.1	Section 1.4	2
2	todo	2
3	todo	2
4	todo	2
4.1	todo	2
4.2	todo	2
4.3	todo	2
4.4	todo	2
4.5	The Dimensions of a Vector Space	2
4.6	Change of basis	4
5	Eigenvalues and eigenvectors	5
5.1	Eigenvectors and eigenvalues	5
5.2	The Characteristic Equation	5
5.3	Diagonalization	6
5.4	Eigenvectors and Linear Transformations	7
5.5	Complex Eigenvalues	8
6	Orthogonality and Least Squares	10
6.1	Inner Product, Length, and Orthogonality	10
6.2	Orthogonal sets	11
6.3	Orthogonal projections in \mathbb{R}^2	12
6.4	The Gram-Schmidt Process	13
6.5	Least-Squares Problems	13
6.6	Linear regression	14
7	Symmetric Matrices and Quadratic Forms	16
7.1	Diagonalization of Symmetric Matrices	16

1 Section 1

1.1 Section 1.4

Relating consistency to linear combinations/spans: Suppose A is an $m \times n$ matrix made of columns $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_n$ and \hat{b} is a vector of length m . Then $A\hat{x} = \hat{b}$ is consistent if and only if $\hat{b} \in \text{span}(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_n)$, i.e., \hat{b} is a linear combination of $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_n$.

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

1. For every $\hat{b} \in \mathbb{R}^m$, the equation $A\hat{x} = \hat{b}$ has a solution.
2. Every vector $\hat{b} \in \mathbb{R}^m$ can be written as a linear combination of the columns of A .
3. $\text{span}_{\mathbb{R}}(\text{columns of } A) = \mathbb{R}^m$.
4. A has a pivot in every row.

2 todo

3 todo

4 todo

4.1 todo

4.2 todo

4.3 todo

4.4 todo

4.5 The Dimensions of a Vector Space

Goldilocks

Theorem: If V is a vector space with basis $B = \{\vec{b}_1, \dots, \vec{b}_n\}$, then any set of vectors in V with more than n elements must be linearly dependent. In particular, any linearly independent set in V has at most n vectors in it.

Theorem: If V is a vector space with basis $B = \{\vec{b}_1, \dots, \vec{b}_n\}$, then any set of vectors in V with less than n elements doesn't span V .

Theorem: If V is a vector space with basis $B = \{\vec{b}_1, \dots, \vec{b}_n\}$, then every basis of V also has exactly n vectors in it.

Definition: If V has a finite spanning set, V is **finite-dimensional**. The size of any basis is called the **dimension** of V , written $\dim(V)$.

Define $\dim(\{\vec{0}\}) = 0$. If not spanned by finite set, is infinite-dimensional.

Dimensions of various spaces

- \mathbb{R}^n has dimension n
- \mathbb{P}^n has dimension $n + 1$ (basis is $\{1, t^1, t^2, \dots, t^n\}$)
- Complex numbers have dimension 2 (because they're $\text{span}(1, i)$)

Theorem: Let V be a finite-dimensional vector space and let $H \subseteq V$ be a vector subspace.

1. Any linearly independent set in H can be expanded to a basis of H .
2. H is finite-dimensional and $\dim H \leq \dim V$.

Theorem: Let V be a finite-dimensional vector space, $\dim V = n \geq 1$.

1. Any linearly independent set of n vectors in V is automatically a basis of V .
2. Any spanning set of n vectors in V is automatically a basis of V .

Dimensions of stuff

1. $\dim(\text{Col}(A)) = \#$ pivot columns of A
2. $\dim(\text{Null}(A)) = \#$ non pivot columns of A
3. $\dim(\text{Row}(A)) = \dim(\text{Col}(A)) = \#$ pivot rows of A

Definition: The **rank** of an $m \times n$ matrix A is $\dim(\text{Col}(A)) = \dim(\text{Row}(A)) = \#$ pivots in A .

Additions to Invertible Matrix Theorem

1. $\text{rank}(A) = n$
2. $\text{Col}(A) = \mathbb{R}^n$
3. $\text{Null}(A) = \{\vec{0}\}$

Fact: Let V, W be finite-dimensional vector spaces, let $T: V \mapsto W$ be a linear transformation.

Then $\dim(\text{Ker}(T)) + \dim(\text{Im}(T)) = \dim W$

(analogous to dimension of null space + col space equaling # cols)

Fact: Row ops don't change relations of linear dependence between columns.
But they do change relations of linear dependence between *rows*.

4.6 Change of basis

If you have a vector \vec{x} in some basis B and you want to find its coordinates in some other basis B' , you can use $[\vec{x}]_{B'} = \left[\begin{bmatrix} \vec{b}_1 \end{bmatrix}_B \quad \begin{bmatrix} \vec{b}_2 \end{bmatrix}_B \right] [\vec{x}]_B$

Change-of-coordinates matrix

The above works in general:

If $B = \{\vec{b}_1, \dots, \vec{b}_n\}$ and $B' = \{\vec{b}'_1, \dots, \vec{b}'_n\}$ are bases of some vector space, then there is a unique $n \times n$ matrix $P_{B' \leftarrow B}$ such that $[\vec{x}]_{B'} = P_{B' \leftarrow B} \cdot [\vec{x}]_B$

where $P_{B' \leftarrow B} = \left[\begin{bmatrix} \vec{b}_1 \end{bmatrix}_{B'} \dots \begin{bmatrix} \vec{b}_n \end{bmatrix}_{B'} \right]$

This $P_{B' \leftarrow B}$ is called the **change-of-coordinates matrix** from B to B' .

Columns of $P_{B' \leftarrow B}$ are coordinate vectors of a basis so they're also linearly independent, so by Invertible Matrix Theorem, $P_{B' \leftarrow B}$ is invertible. So

$$[\vec{x}]_B = \left(P_{B' \leftarrow B} \right)^{-1} \cdot [\vec{x}]_{B'}$$

$$\left(P_{B' \leftarrow B} \right)^{-1} = P_{B \leftarrow B'}$$

Finding change-of-coordinates matrix

Let $B = \{\vec{b}_1, \dots, \vec{b}_n\}$ and $B' = \{\vec{b}'_1, \dots, \vec{b}'_n\}$ be 2 bases of \mathbb{R}^n . To get $P_{B' \leftarrow B}$, make augmented matrix and row reduce:

$$\left[\begin{array}{ccc|ccc} \vec{b}'_1 & \vec{b}_1 & \dots & \vec{b}'_n & \vec{b}_1 & \vec{b}_2 & \dots & \vec{b}_n \end{array} \right] \sim \left[I_n \mid P_{B' \leftarrow B} \right]$$

Fact: There's a 1-1 correspondence between bases and invertible matrices.

5 Eigenvalues and eigenvectors

5.1 Eigenvectors and eigenvalues

Some linear transformations only stretch certain vectors along lines, and finding those lines can be helpful. An eigenvector is just a vector that gets scaled after a linear transformation.

Definition: Let A be an $n \times n$ matrix. Let $\lambda \in \mathbb{R}$. A nonzero vector \vec{v} is called an **eigenvector** of A with **eigenvalue** λ if $A\vec{v} = \lambda\vec{v}$

To check if something is an eigenvector, just multiply $A\vec{v}$, then check if the result is a multiple of \vec{v} .

To find eigenvectors, can take advantage of

$$A\vec{v} = \lambda\vec{v} \Leftrightarrow A\vec{v} - \lambda\vec{v} = \vec{0} \Leftrightarrow (A - \lambda I)\vec{v} = \vec{0} \Leftrightarrow \vec{v} \in \text{Null}(A - \lambda I)$$

So **eigenspace** $E_\lambda = \text{Null}(A - \lambda I) = \{\vec{v} \in \mathbb{R}^n \mid A\vec{v} = \lambda\vec{v}\}$ is a vector subspace of \mathbb{R}^n .

So we know how to find all eigenvectors with a given eigenvalue.

Theorem: If A is upper- or lower-triangular, then the eigenvalues of A are its diagonal entries.

Theorem: Eigenvectors with distinct eigenvalues are linearly independent. That is, if A is an $n \times n$ matrix and $\vec{v}_1, \dots, \vec{v}_k$ are eigenvectors for A with distinct eigenvalues $\lambda_1, \dots, \lambda_k$, then $\{\vec{v}_1, \dots, \vec{v}_k\}$ is a linearly independent set in \mathbb{R}^n .

5.2 The Characteristic Equation

Goal: Finding all eigenvalues.

- We want the eigenspace to not be trivial ($\{\vec{0}\}$).
- For $E_\lambda = \text{Null}(A - \lambda I)$ to not be trivial, $A - \lambda I$ must be singular (not invertible).
- So $\det(A - \lambda I) = 0$.
- Generally, $\det(A - \lambda I)$ is a polynomial, and the roots of that polynomial are exactly the eigenvalues of A .
- **Definition:** We call $\det(A - \lambda I) = P_A(\lambda)$ the **characteristic polynomial** of A .
- The equation $P_A(\lambda) = 0$ is called the **characteristic equation** of A .

Algebraic multiplicity

If the characteristic polynomial looks like $(1 - \lambda)(2 - \lambda)^2(4 - \lambda) = 0$, the root 2 has multiplicity 2 because it's squared. The **algebraic multiplicity** of an eigenvalue is the power its part of the polynomial is raised to in the characteristic polynomial.

If 0 is one of the eigenvalues, A is singular (not invertible).

If an eigenvalue has multiplicity 1, its associated eigenspace has dimension 1.

Similar matrices

Problems with the methods above:

- Computing $P_A(\lambda)$ is inefficient
- Finding its roots is also inefficient

Similar matrices are a better way. Suppose A and B are $n \times n$ matrices, and P is invertible, such that $A = PBP^{-1}$. Then we say A and B are **similar** matrices (similarity is an equivalence relation).

Theorem: Any 2 similar matrices have the same characteristic polynomial and therefore the same eigenvalues with the same multiplicities (the converse is not true, two matrices can have the same characteristic polynomials without being similar).

Main idea: if A is similar to some triangular matrix B , then the eigenvalues of A are just the diagonal entries of B .

Similar matrices are not necessarily row-equivalent (and the other way around) because row operations can change eigenvalues.

Geometric multiplicity and taking powers

The dimension of the eigenspace for some eigenvalue is the **geometric multiplicity** of that eigenvalue.

If two matrices A and D are similar, then A^n and D^n are also similar.

5.3 Diagonalization

Definition: A square matrix A is **diagonalizable** if A is similar to a diagonal matrix, i.e., $A = PDP^{-1}$ for some diagonal matrix D and invertible matrix P .

One nice thing about diagonal matrices is that it's easy to square/cube/take high powers of them. A^n is just all the individual elements raised to n .

This can be extended to *diagonalizable* matrices too: if $A = PDP^{-1}$ where D is a diagonal matrix and A is a diagonalizable matrix, then $A^n = PD^nP^{-1}$.

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

In this case, $A = PDP^{-1}$ where D is a diagonal matrix whose diagonal is $\lambda_1, \lambda_2, \dots, \lambda_n$ where each λ_i is an eigenvalue of A and $P = [\vec{v}_1 \dots \vec{v}_n]$ with $A\vec{v}_i = \lambda_i\vec{v}_i$, which is your basis of eigenvectors.

Algebraic and geometric multiplicities

Fact: If $P_A(\lambda) = (\lambda - a_1)^{r_1}(\lambda - a_2)^{r_2} \dots (\lambda - a_k)^{r_k}$, then

r_i is the **algebraic multiplicity** of a_i

and $\dim(E_{a_i}) = \dim(\text{Null}(A - a_i I))$ is the **geometric multiplicity** of a_i .

$1 \leq \dim(E_{a_i}) \leq r_i$ so geometric multiplicity \leq algebraic multiplicity

So if A is $n \times n$ with n distinct eigenvalues, then $P_A(\lambda) = (\lambda - a_1)(\lambda - a_2) \dots (\lambda - a_n)$

so $\dim(E_{a_i}) = 1$ for each i .

Thus, choose an eigenvector $\vec{v}_i \in E_{a_i}$ ($A\vec{v}_i = a_i\vec{v}_i$)

then $\{\vec{v}_1, \dots, \vec{v}_n\}$ form a basis of eigenvectors, so A is diagonalizable.

Theorem: If A is $n \times n$ with n distinct eigenvalues, then A is diagonalizable (converse is not true).

How to tell if two matrices aren't similar

Check if algebraic and geometric multiplicities are the same.

Can look at Jordan canonical form (don't need to know for exam or anything)

5.4 Eigenvectors and Linear Transformations

If there's a linear transformation $T: V \mapsto W$ and B is a basis for V and C is a basis for W , then if you want a matrix that does what T does (turns coordinate vectors for V relative to B into coordinate vectors for W relative to C), then that matrix is

$M_T^{B,C} = \begin{bmatrix} [T(\vec{b}_1)]_C & \dots & [T(\vec{b}_n)]_C \end{bmatrix}$ where the \vec{b}_i s are the vectors in the standard basis for B .

If $T: V \mapsto V$ and you're using the same basis for both the input and output, then $M_T^{B,B}$ can also be called $[T]_B$ (matrix for T relative to B) in short.

Similar matrices and change of basis

- Suppose A and D are similar $n \times n$ matrices. Then $A = PDP^{-1}$ for some invertible $n \times n$ matrix P .
- Make a linear transformation $T: \mathbb{R}^n \mapsto \mathbb{R}^n$ that maps \vec{x} to $A\vec{x}$ (just the linear transformation that A represents).
- So $[T]_{\mathcal{E}} = A$
- Say $P = \begin{bmatrix} \vec{b}_1 & \dots & \vec{b}_n \end{bmatrix}$ and let $B = \{\vec{b}_1, \dots, \vec{b}_n\}$ (B is a basis of \mathbb{R}^n)
- **Fact:** Then $[T]_B = D$
- So A and D are representations of the same linear transformation, but relative to different fixed bases.

Definition: Let V be a finite-dimensional vector space and $T: V \mapsto V$
A nonzero vector $\vec{v} \in V$ is an **eigenvector** of T with eigenvalue λ if $T(\vec{v}) = \lambda\vec{v}$

So $\vec{v} \in \ker(T - \lambda I) = E_\lambda$ (called **eigenspace**, a vector subspace of V)

If V has a basis B relative to which the matrix for T is diagonal, say $B = \{\vec{b}_1, \dots, \vec{b}_n\}$,
then $[T]_B = \text{diag}([\lambda_1 \dots \lambda_n])$

So $T(\vec{b}_i) = \lambda_i \vec{b}_i$, meaning that B consisted of eigenvectors in the first place.

So if you're picking a basis, make sure to pick a basis made of eigenvectors.

5.5 Complex Eigenvalues

The nice thing about complex eigenvalues is that you're guaranteed to have all n roots/eigenvalues if your characteristic polynomial has degree n . If you want to allow complex eigenvalues, you only need to allow complex entries inside the matrix and its eigenvectors.

Iterates of a vector

If you have a linear transformation $A: \mathbb{R}^2 \mapsto \mathbb{R}^2$, then you can take the **iterates** to understand it better: $A\vec{x}$, $A(A\vec{x})$, $A(A(A\vec{x}))$, ..., $A^k\vec{x}$.

Some notation:

If you have a $z \in \mathbb{C}$ and $z = a + bi$, then $a = \text{Re}(z)$ and $b = \text{Im}(z)$

Similarly, if a vector $\vec{v} \in \mathbb{C}^n$ and $\vec{v} = \begin{bmatrix} a_1 + b_1 i \\ \vdots \\ a_n + b_n i \end{bmatrix}$, then $\text{Re}(\vec{v}) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ and $\text{Im}(\vec{v}) = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$

Definition: The **complex conjugate** of $z = a + bi$ is $\bar{z} = a - bi$

For $\vec{v} = \begin{bmatrix} a_1 + b_1 i \\ \vdots \\ a_n + b_n i \end{bmatrix}$, just take complex conjugate of each entry: $\bar{\vec{v}} = \begin{bmatrix} a_1 - b_1 i \\ \vdots \\ a_n - b_n i \end{bmatrix}$

When A is an $n \times n$ real matrix, $P_A(\lambda)$ has real coefficients.

Fact: The complex roots of $P_A(\lambda)$ come in conjugate pairs (if you have one complex root, its complex conjugate is also a root of $P_A(\lambda)$)

This is important because if $\lambda \in \mathbb{C}$ is an eigenvalue of some matrix A , then $\bar{\lambda}$ (its complex conjugate) is also an eigenvalue of A .

Imagining the complex numbers as 2×2 matrices

If $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, then $A^2 = -I$, so A is kinda like i

If $B = I$, then $B^2 = I$, so B is kinda like 1

You could use these two matrices to make a 2D vector space isomorphic to \mathbb{C} :

$$\left\{ a \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + b \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \mid a, b \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \mid a, b \in \mathbb{R} \right\}$$

This vector space maintains not only the additive structure of \mathbb{C} but also the multiplicative structure, so you can think of complex numbers as these 2×2 matrices instead.

You can think of complex numbers as points on a 2D plane in terms of polar coordinates (r, θ) (where r is the norm and θ is the angle such that $a = r \cos \theta$ and $b = r \sin \theta$)

So multiplying by a complex number is like scaling by r and rotating it by θ (if you convert it to one of the 2×2 matrices above, it looks like a rotation matrix times r).

Theorem:

Let A be a real 2×2 matrix with complex eigenvalues $\lambda = a \pm bi$, $b > 0$.

Let $\vec{v} \in \mathbb{C}^2$ be an eigenvector with eigenvalue $a - bi$.

Take $P = [\text{Re}(\vec{v}) \quad \text{Im}(\vec{v})]$

Then P is invertible and $P^{-1}AP = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ (A is similar to that complex number)

This sort of thing also happens in higher dimensions but that won't be in this class

6 Orthogonality and Least Squares

6.1 Inner Product, Length, and Orthogonality

Dot product

Definition: For $\vec{u}, \vec{v} \in \mathbb{R}^n$, $\vec{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$, $\vec{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$, define $\vec{u} \cdot \vec{v} = u_1 v_1 + \dots + u_n v_n = \vec{u}^T \vec{v}$

Facts:

- $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$ (commutative)
- $\vec{u} \cdot (\vec{v} + \vec{w}) = \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w}$ (distributive)
- $(c\vec{u}) \cdot \vec{v} = c(\vec{u} \cdot \vec{v})$
- $\vec{u} \cdot \vec{u} \geq 0$ and $\vec{u} \cdot \vec{u} = 0$ only if $\vec{u} = \vec{0}$

Length and Distance

Definition: The **norm** (or length) of a vector is written as $\|\vec{u}\| = \sqrt{\vec{u} \cdot \vec{u}}$

Properties:

- $\|c\vec{u}\| = |c| \cdot \|\vec{u}\|$ (c comes out but with an absolute value around it)
- $\left\| \frac{\vec{w}}{\|\vec{w}\|} \right\| = 1$
- $\|\vec{u} - \vec{v}\|$ is the difference from the tip of \vec{u} to the tip of \vec{v}

Definition: A **unit vector** is any vector \vec{v} such that $\|\vec{v}\| = 1$

We can always find a unit vector pointing in the same direction as some vector \vec{w}

Angle

Dot product also relates to angle: $\vec{u} \cdot \vec{v} = \|\vec{u}\| \cdot \|\vec{v}\| \cos \theta$ (θ is angle between the vectors)

So $\theta = \cos^{-1} \left(\frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \cdot \|\vec{v}\|} \right)$

Special case: Iff $\theta = \frac{\pi}{2}$ (the vectors are perpendicular/orthogonal), then $\vec{u} \cdot \vec{v} = 0$

We declare \vec{u} and \vec{v} to be **orthogonal** if $\vec{u} \cdot \vec{v} = 0$

$\vec{0}$ is orthogonal to everything

Orthogonal complement

Definition: The orthogonal complement of W is $W^\perp = \{\vec{v} \in \mathbb{R}^n \mid (\forall \vec{w} \in W, \vec{v} \cdot \vec{w} = 0)\}$

Facts:

- W^\perp is a vector subspace
- $\vec{v} \in W^\perp$ iff \vec{v} is orthogonal to every vector of some spanning set of W

Theorem: Let A be an $m \times n$ matrix. Then

1. $\text{Row}(A)^\perp = \text{Null}(A)$
2. $\text{Col}(A)^\perp = \text{Null}(A^T)$

6.2 Orthogonal sets

Definition: Say $S = \{\vec{u}_1, \dots, \vec{u}_k\} \subseteq \mathbb{R}^n$ is a set of vectors.

S is an **orthogonal set** if $\vec{u}_i \cdot \vec{u}_j = 0$ whenever $i \neq j$ (every pair of vectors in S is orthogonal)

Theorem: If S is an orthogonal set of nonzero vectors in \mathbb{R}^n , then S is linearly independent.

Orthogonal bases

Using the theorem above, we get that an orthogonal set of nonzero vectors is a basis for its span. This is called an **orthogonal basis**.

Theorem: If $B = \{\vec{u}_1, \dots, \vec{u}_k\}$ is an orthogonal basis of $W = \text{span}\{\vec{u}_1, \dots, \vec{u}_k\} \subseteq \mathbb{R}^n$, then for any $\vec{y} \in W$,

$$[\vec{y}]_B = \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix} \text{ where } c_i = \frac{\vec{y} \cdot \vec{u}_i}{\vec{u}_i \cdot \vec{u}_i}$$

In order to tell if a vector is in the span of an orthogonal basis, you can try to get its coordinates relative to the orthogonal basis, and if it doesn't work out, you know it's not in the span.

Orthogonal projection

Definition: The orthogonal projection of \vec{y} onto \vec{u} is $\hat{y} = \text{proj}_{\vec{u}}(\vec{y}) = \frac{\vec{y} \cdot \vec{u}}{\vec{u} \cdot \vec{u}} \vec{u}$

It has some useful properties:

1. It's a scalar multiple of \vec{u}
2. $\hat{y} \cdot \vec{u} = \vec{y} \cdot \vec{u}$, which means $(\vec{y} - \hat{y}) \cdot \vec{u} = 0$, which means $\vec{y} - \hat{y}$ is orthogonal to \vec{u}
3. \hat{y} is the only scalar with the above two properties

So this breaks \vec{y} into the component that's parallel to \vec{u} (\hat{y}) and the component that's perpendicular to \vec{u} ($\vec{y} - \hat{y}$).

Orthonormal sets

Definition: An orthonormal set is an orthogonal set where the norm of all the vectors is 1, i.e. $\vec{u} \cdot \vec{u} = 1$ for all \vec{u} in that set.

You can get an orthonormal set from an orthogonal set by normalizing the vectors in it.

Fact: An $m \times n$ matrix U has orthonormal columns iff $U^T U = I$

Fact: If U has orthonormal columns, then

- $\|\vec{x}\| = \|U\vec{x}\|$ (U preserves length)
- $\vec{x} \cdot \vec{y} = (U\vec{x}) \cdot (U\vec{y})$ (U preserves angles)
- $U^T = U^{-1}$

Definition: When U is a square matrix whose columns are *orthonormal*, U is called an **orthogonal matrix** (not “orthonormal” matrix, there’s no word for a matrix whose columns are orthogonal but not orthonormal)

6.3 Orthogonal projections in \mathbb{R}^2

Let $W \subseteq \mathbb{R}^n$ be a subspace, say $W = \text{span}\{\vec{u}_1, \dots, \vec{u}_k\}$

Let $\vec{y} \in \mathbb{R}^n$

We want to find a vector $\text{proj}_W(\vec{y})$ such that

1. $\text{proj}_W(\vec{y}) \in W$ and
2. $\vec{y} - \text{proj}_W(\vec{y}) \in W^\perp$

i.e., we want to decompose \vec{y} into two orthogonal components.

What if $\{\vec{u}_1, \dots, \vec{u}_k\}$ were an orthogonal basis of W ?

Define $\text{proj}_W(\vec{y}) = \sum_{i=1}^k \text{proj}_{\vec{u}_i}(\vec{y})$

Fact: This orthogonal projection is unique.

Fact: $\vec{y} = \text{proj}_W(\vec{y}) \Leftrightarrow \vec{y} \in W$

Orthogonal projections as approximations

Theorem:

Let $W \subseteq \mathbb{R}^n$ be a subspace

Let $\vec{y} \in \mathbb{R}^n$

Then $\text{proj}_W(\vec{y})$ is the vector in W that’s closest to \vec{y}

In other words, $\|\vec{y} - \text{proj}_W(\vec{y})\| < \|\vec{y} - \vec{w}\|$ for any vector \vec{w} in W such that $\vec{w} \neq \text{proj}_W(\vec{y})$

If you want to solve a solution $A\vec{x} = \vec{b}$ but it's inconsistent, you can solve $A\vec{x} = \text{proj}_W(\vec{b})$ instead to get the closest thing to a solution.

Orthogonal projections and orthonormal bases

Theorem:

Let $\{\vec{u}_1, \dots, \vec{u}_k\}$ be an orthonormal basis of $W \subseteq \mathbb{R}^n$

1. For any $\vec{y} \in \mathbb{R}^n$, $\text{proj}_W(\vec{y}) = \sum_{i=1}^k (\vec{y} \cdot \vec{u}_i) \vec{u}_i$
2. Set $U = [\vec{u}_1 \ \dots \ \vec{u}_k]$. Then for any $\vec{y} \in \mathbb{R}^n$, $\text{proj}_W(\vec{y}) = UU^T \cdot \vec{y}$

6.4 The Gram-Schmidt Process

Used to iteratively turn a basis into an orthogonal basis.

Suppose you have a basis $\{\vec{x}_1, \dots, \vec{x}_n\}$

Steps:

- $\vec{v}_1 = \vec{x}_1$
- $\vec{v}_2 = \vec{x}_2 - \text{proj}_{\vec{v}_1}(\vec{x}_2)$ (can scale \vec{v}_2 if you want to make it nicer)
- $\vec{v}_3 = \vec{x}_3 - \text{proj}_{\vec{v}_1}(\vec{x}_3) - \text{proj}_{\vec{v}_2}(\vec{x}_3)$
- ...
- $\vec{v}_k = \vec{x}_k - \text{proj}_{\text{span}\{\vec{v}_1, \dots, \vec{v}_{k-1}\}}(\vec{x}_k) = \vec{x}_k - \text{proj}_{\vec{v}_1}(\vec{x}_k) - \dots - \text{proj}_{\vec{v}_{k-1}}(\vec{x}_k)$

The new orthogonal basis is $\{\vec{v}_1, \dots, \vec{v}_n\}$

This produces $\vec{v}_1, \dots, \vec{v}_n$ in such a way that each $\vec{v}_k \in \text{span}\{\vec{x}_1, \dots, \vec{x}_k\}$

QR Factorization

Take those vectors produced by the Gram process and normalize: $\vec{u}_i = \frac{\vec{v}_i}{\|\vec{v}_i\|}$

Set $Q = [\vec{u}_1 \ \dots \ \vec{u}_n]$

Then there is an $n \times n$ upper triangular matrix R with positive diagonal entries (so invertible) such that $A = QR$ (where the columns of A were $\vec{x}_1, \dots, \vec{x}_n$)

6.5 Least-Squares Problems

If an equation $A\vec{x} = \vec{b}$ doesn't have a solution, you want the best approximate solution. Least squares is a way of judging how good a solution is.

We want $\hat{x} \in \mathbb{R}^n$ such that $\|A\hat{x} - \vec{b}\| \leq \|A\vec{x} - \vec{b}\|$ for all $\vec{x} \in \mathbb{R}^n$

Set $\hat{b} = \text{proj}_{\text{Col}(A)}(\vec{b})$ so that $A\hat{x} = \hat{b}$ for at least 1 $\hat{x} \in \mathbb{R}^n$.

Definition: Any such \hat{x} is called a **least squares solution** of $A\vec{x} = \vec{b}$

The **least squares error** is $\|A\hat{x} - \vec{b}\|$

Fact: Iff $A\hat{x} = \hat{b}$, then $A^T\vec{b} = A^TA\hat{x}$

Normal equation for $A\vec{x} = \vec{b}$: $A^TA\vec{x} = A^T\vec{b}$

Summary: To get least squares solutions of $A\vec{x} = \vec{b}$, we solve $A^TA\vec{x} = A^T\vec{b}$ (normal equation)

The set of solutions to the normal equation is the same as the set of least squares solutions.

Fact: $\text{Null}(A^TA) = \text{Null}(A)$

Theorem: For an $m \times n$ matrix A , the following are equivalent:

1. A^TA is invertible
2. The columns of A are linearly independent
3. $A\vec{x} = \vec{b}$ has a unique least squares solution for any $\vec{b} \in \mathbb{R}^m$

A^TA is called the **Gram Matrix**. If A is diagonal, then A^TA is a diagonal matrix.

Fact: If A has orthogonal columns, then the least squares solution is $\text{proj}_{\text{Col}(A)}(\vec{b})$

When A has linearly independent columns, you can do QR factorization to get Q and R . Then the least squares solution to $A\vec{x} = \vec{b}$ is $\hat{x} = R^{-1}Q^T\vec{b}$ and you're done.

6.6 Linear regression

Linear relationships

Motivation: Trying to find line of best for some data $(x_1, y_1), \dots, (x_n, y_n)$. Line of best fit is in the form $y = \beta_0 + \beta_1 x$

When trying to find the line of best fit, we want to choose a line where the sum of the squares of the vertical distance between the points and the line is minimized. Those vertical distances between the points and the line of best fit are called **residuals**.

If the points really were on that line of best fit, they'd satisfy the equation $X\vec{\beta} = \vec{y}$, where

$$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \text{ and } \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

X is called the **design matrix**, \vec{y} is called the **observation vector**.

We find the least squares solution by solving $X^TX\vec{\beta} = X^T\vec{y}$ instead of $X\vec{\beta} = \vec{y}$.

Other regressions

Best-fit question is linear even when model is a curve. For example, if you think some data can be best represented by something in the form $y = \beta_0 + \beta_1x + \beta_2x^2$, then you can set

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \text{ and use } \vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

You can do this with trig functions and stuff too. As long as it's in the form $y = \beta_0f_0(x) + \dots + \beta_kf_k(x)$ (for some known f_0, \dots, f_k , it's still a linear equation in β_0, \dots, β_k). Just set

$$X = \begin{bmatrix} f_0(x_1) & \dots & f_k(x_1) \\ \vdots & \ddots & \vdots \\ f_0(x_n) & \dots & f_k(x_n) \end{bmatrix} \text{ and } \vec{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}$$

Multivariate models

Suppose you now have 2 independent variables u and v and your model needs to look like

$$y = \beta_0 + \beta_1u + \beta_2v + \beta_3uv. \text{ You can still do the same thing, just set } X = \begin{bmatrix} 1 & u_1 & v_1 & u_1v_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & u_n & v_n & u_nv_n \end{bmatrix}$$

$$\text{and } \vec{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_4 \end{bmatrix}$$

7 Symmetric Matrices and Quadratic Forms

7.1 Diagonalization of Symmetric Matrices

Spectral Theorem

Definition: Given an $n \times n$ matrix A , the **spectrum** of A is $\{\text{eigenvalues of } A\}$.

Definition: A matrix A is **symmetric** if $A = A^T$ ($a_{ij} = a_{ji}$, if $A = (a_{ij})$)

Fact: (bit unrelated) Symmetric matrices have real eigenvalues.

Theorem: Let A be symmetric. Then eigenvectors for A with distinct eigenvalues are orthogonal (if it weren't symmetric, they'd merely be linearly independent).

One nice thing is that now if you diagonalize A into PDP^{-1} , then P just needs to be scaled to become an orthogonal matrix Q (you can keep the same D , so now $A = QDQ^{-1} = QDQ^T$).

Definition: If $A = QDQ^T$ for a diagonal matrix D and an orthogonal matrix Q , then we say that A is **orthogonally diagonalizable**.

Theorem (Spectral Theorem): An $n \times n$ real matrix is orthogonally diagonalizable if and only if it is symmetric.

Spectral Decomposition

If you want to multiply some matrix A with some matrix B , you can multiply the columns of A by the rows of B instead of the rows of A with the columns of B . So if A 's columns are \vec{a}_i and B 's rows are \vec{b}_i , then $AB = \vec{a}_1\vec{b}_1^T + \dots + \vec{a}_n\vec{b}_n^T$. Each of these $\vec{a}_i\vec{b}_i^T$ are rank 1 matrices.

If you have a symmetric matrix A that has eigenvectors $\vec{u}_1, \dots, \vec{u}_n$ with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$, you can factor it into QDQ^T , where $Q = [\vec{u}_1 \ \dots \ \vec{u}_n]$ and $D = \begin{bmatrix} \lambda_1 & \dots & 0 \\ 0 & \ddots & 0 \\ 0 & \dots & \lambda_n \end{bmatrix}$.

Using that thing above, $QDQ^T = \lambda_1\vec{u}_1\vec{u}_1^T + \dots + \lambda_n\vec{u}_n\vec{u}_n^T$.

This is called the **spectral decomposition** of A .