

ECE367 Cheatsheet

Hanhee Lee

September 16, 2024

Contents

1	Vectors, Norms, Inner Products (Ch. 2.1-2.2)	3
1.1	Linear transformation	3
1.1.1	Matrix representation of a linear transformation	3
1.2	Vectors	3
1.3	Vector spaces	3
1.3.1	How to prove or disprove a vector space?	4
1.4	Subspace	5
1.5	Span	5
1.5.1	How to draw the span?	5
1.6	Linear independent (LI) set	6
1.6.1	How to determine if a set is linearly independent	6
1.7	Basis	6
1.7.1	Dimension	7
1.8	Norms (Notion of distance)	7
1.8.1	Norm balls	8
1.8.2	Motivation for Norms	8
1.8.3	Distance metric	8
1.9	Inner product (Notion of angle)	9
1.9.1	Examples of inner products	9
1.9.2	Connection of inner product to angle	9
1.9.3	Cauchy-Schwartz inequality and its generalization	10
1.9.4	Inner product induces a norm	10
1.10	Orthogonal decomposition	10
1.10.1	Mutually orthogonal	10
1.10.2	Orthonormal basis	11
1.10.3	Orthogonal	11
1.10.4	Orthogonal complement	11
2	Orthogonal Decomposition, Projecting onto Subspaces, Gram-Schmidt, QR Decomposition, Hyperplanes and Half-Spaces (Ch. 2.2-2.3)	11
2.1	Projection onto subspaces	11
2.1.1	Basic problem	12
2.1.2	Projection onto a 1D subspace	13
2.1.3	Projection onto an n dimensional space	13
2.1.4	Application of projections: Fourier series	15
2.2	Gram-Schmidt and QR decomposition	15
2.2.1	What if the set of basis vectors is not orthonormal?	15
2.2.2	Gram-Schmidt Procedure	17
2.2.3	QR decomposition	19
2.2.4	Eigen-composition	19
2.2.5	Singular value decomposition	19
2.3	Hyperplanes and half-spaces	19
2.4	Non-euclidean projection	19

3 Non-Euclidean Projection, Projection onto Affine Sets, Functions, Gradients and Hessians (Ch. 2.3-2.4)	19
3.1 Projection onto affine sets	19
3.2 Functions	19
3.3 Gradients	19
3.4 Hessians	19
4 Matrices, Range, Null Space, Eigenvalues, Eigenvectors, Matrix Diagonalization (Ch. 3.1-3.5)	19
4.1 Matrices	19
4.2 Range	19
4.3 Null Space	19
4.4 Eigenvalues and eigenvectors	19
4.5 Matrices diagonalization	19
5 Symmetric Matrices, Orthogonal Matrices, Spectral Decomposition, Positive Semidefinite Matrices, Ellipsoids (Ch. 4.1-4.4)	19
5.1 Symmetric matrices	19
5.2 Orthogonal matrices	19
5.3 Spectral decomposition	19
5.4 Positive semidefinite matrices	19
5.5 Ellipsoids	19
6 Singular Value Decomposition, Principal Component Analysis (Ch. 5.1, 5.3.2)	19
6.1 Singular value decomposition	19
6.2 Principle component analysis	19
7 Interpretations of SVD, Low-Rank Approximation (Ch. 5.2-5.3.1)	19
7.1 Interpretation of SVD	19
7.2 Low-rank approximation	19
8 Least Squares, Overdetermined and Underdetermined Linear Equations (Ch. 6.1-6.4)	19
8.1 Least squares	19
8.2 Overdetermined linear equation	19
8.3 Underdetermined linear equation	19
9 Regularized Least-Squares, Convex Sets and Convex Functions (Ch. 6.7.3, 8.1-8.4)	19
9.1 Regularized least-squares	19
9.2 Convex sets and convex functions	19
10 Lagrangian Method for Constrained Optimization, Linear Programming and Quadratic Programming (Ch. 8.5, 9.1-9.6)	19
10.1 Lagrangian method for constrained optimization	19
10.2 Linear programming and quadratic programming	19
11 Numerical Algorithms for Unconstrained and Constrained Optimization (Ch. 12.1-12.3)	19
11.1 Numerical algorithms for unconstrained optimization	19
11.2 Numerical algorithms for constrained optimization	19

List of Figures

1 Vector addition and scalar multiplication.	4
2 Norm balls of different p values.	8
3 Ordering of the vector spaces.	10
4 Drawing any x.	11
5 Error vector being perp. to S.	12
6 Visual representation of the projection problem.	13
7 Generalization of projection.	14
8 Periodic triangle function.	15
9 Not orthogonal, but similar to projection with orthonormal basis.	16
10 Gram-Schmidt Process for 2D.	17

List of Tables

1 Vectors, Norms, Inner Products (Ch. 2.1-2.2)

1.1 Linear transformation

Definition: $T : X \rightarrow Y$ that satisfies

1. **Additivity:** $T(x_1 + x_2) = T(x_1) + T(x_2)$
 2. **Homogeneity:** $T(\alpha x) = \alpha T(x)$
- **Note:** Linear algebra is the study of linear transformations over vector spaces.

1.1.1 Matrix representation of a linear transformation

Definition: Let \mathcal{V} and \mathcal{W} be vector spaces. Let $T : \mathcal{V} \rightarrow \mathcal{W}$ be a linear transformation. When $\mathcal{V} = \mathbb{R}^n$ (or \mathbb{C}^n) and $\mathcal{W} = \mathbb{R}^m$ (or \mathbb{C}^m), then T can be uniquely represented as a matrix $A \in \mathbb{R}^{m \times n}$ such that:

$$T(\mathbf{x}) = A\mathbf{x}$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- **Key:** Any linear transformation is a matrix multiplication. Any matrix multiplication is a linear transformation.

1.2 Vectors

Definition: Ordered collection of numbers, where $x_i \in \mathbb{R}$ or \mathbb{C}

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{x}^T = [x_1 \quad x_2 \quad \cdots \quad x_n].$$

- n : Dimension of \mathbf{x}
- \mathbf{x} : Column vector
- \mathbf{x}^T : Transpose of \mathbf{x} (row vector)
- T : Transpose
- x_i : i -th element of \mathbf{x} .

1.3 Vector spaces

Definition: A vector space over a field \mathbb{F} (e.g. \mathbb{R}/\mathbb{C}) consists of:

1. A set of vectors \mathcal{V}
 2. A vector addition operator $+$: $\mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V}$ s.t. $\forall x, y \in \mathcal{V} \rightarrow x + y \in \mathcal{V}$ (i.e. closed under VA)
 3. A scalar multiplication operator \cdot : $\mathbb{F} \times \mathcal{V} \rightarrow \mathcal{V}$ s.t. $\forall \alpha \in \mathbb{F}, \forall x \in \mathcal{V} \rightarrow \alpha x \in \mathcal{V}$ (i.e. closed under SM)
- \times is not scalar multiplication.

For $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{V}$ and $\alpha, \beta \in \mathbb{F}$. The following properties are satisfied:

- **Vector addition** satisfies (i.e., Abelian group):
 1. **Commutativity:** $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
 2. **Associativity:** $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$.

- 3. **Additive identity:** $\exists \mathbf{0} \in \mathcal{V}$ s.t. $\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$.
- 4. **Additive inverse:** $\forall \mathbf{x}, \exists \mathbf{y}$ s.t. $\mathbf{x} + \mathbf{y} = \mathbf{0}$ (i.e. $\mathbf{y} = -\mathbf{x}$).
- **Scalar multiplication** satisfies:
 - 1. **Associativity:** $\alpha \cdot (\beta \cdot \mathbf{x}) = (\alpha \cdot \beta) \cdot \mathbf{x}$.
 - 2. **Multiplicative Identity:** $\exists 1 \in \mathbb{F}$ s.t. $1 \cdot \mathbf{x} = \mathbf{x}$.
 - 3. **Right Distributivity:** $\alpha \cdot (\mathbf{x} + \mathbf{y}) = \alpha \cdot \mathbf{x} + \alpha \cdot \mathbf{y}$.
 - 4. **Left Distributivity:** $(\alpha + \beta) \cdot \mathbf{x} = \alpha \cdot \mathbf{x} + \beta \cdot \mathbf{x}$.

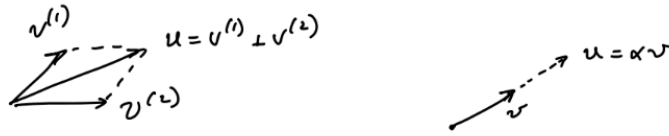


Figure 1: Vector addition and scalar multiplication.

1.3.1 How to prove or disprove a vector space?

Process:

Prove:

1. Prove that \mathcal{V} is closed under VA and SM.
2. Prove all the properties under VA and SM.

Disprove:

1. Disprove one of the properties or that it isn't closed under VA and SM.

Warning: If standard addition and multiplication then, closed under VA and SM properties is enough to prove it's a vector space.

Example:

- Let $\mathcal{V} = \mathbb{R}^n$ and $\mathbb{F} = \mathbb{R}$: This represents vectors of dimension n where each element belongs to \mathbb{R} .

$$\mathcal{V} = \mathbb{R}^n = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} : x_i \in \mathbb{R} \right\}$$

For $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n$:

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

For $\alpha, \beta \in \mathbb{R}$:

$$\alpha \cdot \mathbf{x} = \begin{bmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{bmatrix}$$

- Let $\mathcal{V} = \mathbb{C}^n$ and $\mathbb{F} = \mathbb{C}$: This represents vectors of dimension n with complex components.

$$\mathcal{V} = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} : x_i \in \mathbb{C} \right\}$$

\mathcal{V} is a vector space over \mathbb{C} under element-wise addition and scalar multiplication.

- Let $\mathcal{V} = \{\text{set of all continuous functions } f : \mathbb{R} \rightarrow \mathbb{R}^n\}$ and $\mathbb{F} = \mathbb{R}$:
Let $f_1, f_2 \in \mathcal{V}$, and for $t \in \mathbb{R}$:

$$(f_1 + f_2)(t) = f_1(t) + f_2(t) \Rightarrow f_1 + f_2 \in \mathcal{V}$$

For $\alpha \in \mathbb{R}$:

$$(\alpha f)(t) = \alpha f(t) \Rightarrow \alpha f \in \mathcal{V}$$

- f is the vector, $\mathbb{R} \rightarrow \mathbb{R}^n$ is the input-output relationship. For 2D, $f(x) = [x_1, x_2]^T$, where x is the input, the vector is the output in 2D, and the vector is f .
- Let $\mathcal{V} = \mathcal{P}_n$, the set of all polynomials with real coefficients and degree $\leq n$:

$$\mathcal{V} = \mathcal{P}_n = \{p(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n : a_0, a_1, \dots, a_n \in \mathbb{R}\}$$

\mathcal{V} is a vector space over \mathbb{R} under standard addition and scalar multiplication.

1.4 Subspace

Definition: A **subspace** is a subset of a vector space \mathcal{V} that is a vector space by itself.

- Test:** To check whether a subset is a subspace, check that it is closed under VA & SM.

Example:

- Let $\mathcal{V} = \mathbb{R}^3$, and consider the set:

$$S = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

This set S is a subspace of \mathbb{R}^3 .

- Let $\mathcal{V} = \mathbb{R}^3$, and consider the set:

$$S = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

This set S is **not** a subspace of \mathbb{R}^3 because adding two vectors will make the last component 2.

- Let $\mathcal{V} = \mathbb{R}^n$, and consider the set:

$$S = \{\mathbf{0}\}$$

This set S is a subspace of \mathbb{R}^n .

1.5 Span

Definition: Given a finite set of vectors $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ in the same vector space \mathcal{V} over some field \mathbb{F} then,

$$\text{Span}(S) = \left\{ \sum_{i=1}^m \alpha_i \mathbf{v}_i \mid \alpha_i \in \mathbb{F} \right\}$$

- Note:** $\text{Span}(S)$ is always a subspace of V .

1.5.1 How to draw the span?

Process:

- Identify the vectors.
- Plot the vectors: Plot each vector on a coordinate plane starting at the origin.
- Draw the span: Extend the vectors in both directions to show the line or plane formed by their span. If they span the entire plane, draw dashed lines extending their direction.

Example:

- Let $S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\}$:

$$\text{span}(S) = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

This set $\text{span}(S)$ forms a plane in \mathbb{R}^3 . The vectors span the xy-plane with the z-coordinate fixed at zero.

- Let $S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix} \right\}$:

$$\text{span}(S) = \left\{ x \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} : x \in \mathbb{R} \right\}$$

In this case, $\text{span}(S)$ is a line in \mathbb{R}^3 along the x-axis with y and z coordinates fixed at zero.

1.6 Linear independent (LI) set

Definition: A set of vectors $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is LI if no vector in S can be written as a LC of other vectors in S .

In other words, the only α_i 's that makes $\sum_{i=1}^m \alpha_i \mathbf{v}_i = \mathbf{0}$ is $\alpha_i = 0, \forall i$.

- If $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is LI, then $\forall \mathbf{u} \in \text{span}(S)$, there is a **unique** set of α_i 's s.t. $\mathbf{u} = \sum_{i=1}^k \alpha_i \mathbf{v}_i$ (i.e. there is no redundancies in representation)
 - Coordinates:** $\{\alpha_1, \alpha_2, \dots, \alpha_k\}$ of \mathbf{u} w.r.t. S .
- If S is linearly dependent, then one of the vectors can be written as a LC of the other vectors. In this case, we can remove that vector and continue this process until the remaining set is LI.
 - Note:** Such an irreducible linearly independent set is called a **basis** of $\text{span}(S)$.

1.6.1 How to determine if a set is linearly independent

Process:

- Write a linear combination with coefficients $\alpha_1, \dots, \alpha_k$.
- Set the linear combination equal to 0.
- Solve for $\alpha_1, \dots, \alpha_k$ by solving the set of equations (i.e. each component is one equation).
- If $\alpha_1 = \dots = \alpha_k = 0$, then it is linearly independent.
- Else, linearly dependent by finding a counter example, where the linear combination is 0 for $\alpha_1, \dots, \alpha_k$ not all equal to 0.

1.7 Basis

Definition: A set of vectors B is a basis of a vector space \mathcal{V} if

- B is LI
- $\text{Span}(B) = \mathcal{V}$

Example: What is the standard basis for $\mathcal{V} = \mathbb{R}^n$?

$$\mathbf{e}^{(1)} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}^{(2)} = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad \mathbf{e}^{(n)} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \quad \text{for } \mathbb{R}^n$$

If $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$, then:

$$\mathbf{x} = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + \dots + x_n \mathbf{e}_n$$

1.7.1 Dimension

Definition: The dimension is the number of basis vectors.

- **Note:** Basis is not unique. But $\dim(\mathcal{V})$ is well-defined.

Example:

- $\dim \left(\text{span} \left(\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\} \right) \right) = 2$
- $\dim \left(\text{span} \left(\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\} \right) \right) = 1$
- $\dim(\{\mathbf{0}\}) = 0$
- The dimension for $\mathcal{V} = \mathbb{R}^n$ of the standard basis is n

1.8 Norms (Notion of distance)

Definition: Let \mathcal{V} be a vector space over \mathbb{R} or \mathbb{C} . A norm is a function $\|\cdot\|: \mathcal{V} \rightarrow \mathbb{R}$ that satisfies

1. **Non-negativity:** $\|\mathbf{x}\| \geq 0$, $\forall \mathbf{x} \in \mathcal{V}$, and $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = \mathbf{0}$
2. **Homogeneity:** $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\| \quad \forall \mathbf{x} \in \mathcal{V}, \alpha \in \mathbb{F}$
3. **Triangle inequality:** $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|, \forall \mathbf{x}, \mathbf{y} \in \mathcal{V}$ (triangular inequality)

Example: ℓ_p norms:

$$\|\mathbf{x}\|_p \equiv \left(\sum_{k=1}^n |x_k|^p \right)^{1/p}, \quad 1 \leq p < \infty.$$

- **Note:** For $p < 1$, triangular inequality doesn't hold.
1. **Sum-of-absolute-values length** $p = 1$: $\|\mathbf{x}\|_1 \equiv \sum_{k=1}^n |x_k|$
 2. **Euclidean length** $p = 2$: $\|\mathbf{x}\|_2 \equiv \sqrt{\sum_{k=1}^n x_k^2}$
 3. **Max absolute value norm** $p = \infty$: $\|\mathbf{x}\|_\infty \equiv \max_{k=1, \dots, n} |x_k|$
 - Largest term will dominate as if we common factor out the largest term, each of the other terms will go to 0 as noted in the lp norm.
 4. **Cardinality** $p = 0$: The number of non-zero vectors in \mathbf{x} is

$$\|\mathbf{x}\|_0 = \text{card}(\mathbf{x}) \equiv \sum_{k=1}^n \mathbb{I}(x_k \neq 0), \quad \text{where} \quad \mathbb{I}(x_k \neq 0) \equiv \begin{cases} 1 & \text{if } x_k \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

- **Key:** Not a norm since $\|\alpha \mathbf{x}\|_0 = \|\mathbf{x}\|_0 \neq |\alpha| \cdot \|\mathbf{x}\|_0$ (e.g. if $\alpha = 2$ then this would double the count of number of non-zero vectors for the RS)

1.8.1 Norm balls

Definition: The set of all vectors with ℓ_p norm less than or equal to one,

$$B_p = \{\mathbf{x} : \|\mathbf{x}\|_p \leq 1\} \quad (1)$$

Example: For 2D, the norm balls are as follows:

- $\ell_2 : B_2 = \left\{ \mathbf{x} \mid \sqrt{x_1^2 + x_2^2} \leq 1 \right\}$
- $\ell_1 : B_1 = \{ \mathbf{x} \mid |x_1| + |x_2| \leq 1 \}$
- $\ell_\infty : B_\infty = \{ \mathbf{x} \mid \max |x_i| \leq 1 \text{ or } |x_1| \leq 1, |x_2| \leq 1 \}$
- $\ell_0 : B_0 = \{ \mathbf{x} \mid \text{card}(\mathbf{x}) \leq 1 \}$

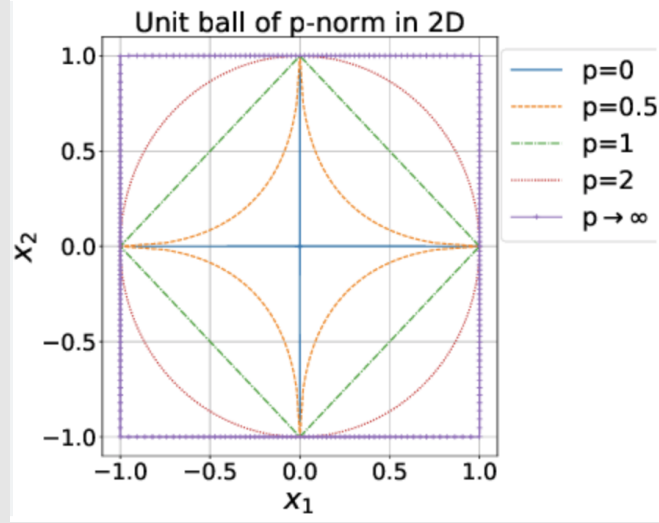


Figure 2: Norm balls of different p values.

1.8.2 Motivation for Norms

Example: In optimization problems, different norms are used to achieve various goals. Suppose we are trying to solve an optimal control problem, where $x = (x_1, \dots, x_n)$ are some action variables.

- $\min \|\mathbf{x}\|_2^2 = x_1^2 + \dots + x_n^2$ (i.e. minimizing the total energy (power) in \mathbf{x})
- $\min \|\mathbf{x}\|_\infty$ (i.e. minimizing the peak energy in \mathbf{x}).
- $\min \|\mathbf{x}\|_1$ (i.e. minimizing the sum of action variables).
- $\min \|\mathbf{x}\|_0$ (i.e. find sparse solution)

1.8.3 Distance metric

Definition: A norm induces a distance metric between two vectors x and y in \mathbb{V} as

$$d(x, y) = \|x - y\|$$

- **Note:** The ℓ_2 -norm induces the Euclidean distance

$$\|x - y\|_2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

1.9 Inner product (Notion of angle)

Definition: An inner product on a vector space \mathcal{V} is a function $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{F}$ such that:

1. **Positive definiteness:** $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0 \ \forall \mathbf{x} \in \mathcal{V}$ and $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ iff $\mathbf{x} = 0$
2. **Conjugate Symmetry:** $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$
 - $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ in \mathbb{R}^n
 - $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$ in \mathbb{C}^n .
3. **Linearity in first argument:** $\langle \alpha \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \alpha \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle \quad \forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{V}, \alpha \in \mathbb{F}$

Example: How to use the properties of inner products?

$$\begin{aligned} \langle x, \alpha y + z \rangle &\stackrel{(2)}{=} \overline{\langle \alpha y + z, x \rangle} \\ &\stackrel{(3)}{=} \overline{\alpha \langle y, x \rangle + \langle z, x \rangle} \quad \text{also by conjugate prop.} \\ &\stackrel{(2)}{=} \overline{\alpha} \overline{\langle y, x \rangle} + \overline{\langle z, x \rangle} \\ &\stackrel{(2)}{=} \overline{\alpha} \langle x, y \rangle + \langle x, z \rangle \end{aligned}$$

1.9.1 Examples of inner products

Example:

- In \mathbb{R}^n (Dot product): $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i = \mathbf{x}^\top \mathbf{y} = \mathbf{y}^\top \mathbf{x}$
 - **Key:** $\langle \mathbf{x}, \mathbf{x} \rangle = \sum_{i=1}^n x_i^2 = \mathbf{x}^\top \mathbf{x} = \|\mathbf{x}\|_2^2$
- In \mathbb{C}^n : $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n \overline{x_i} y_i = \mathbf{y}^H \mathbf{x} = \overline{\mathbf{x}^H \mathbf{y}}$
 - $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \mathbf{x}^H = [\overline{x_1} \quad \cdots \quad \overline{x_n}]$
- $\mathcal{V} = \left\{ f : \mathbb{R} \rightarrow \mathbb{R}; \int_{-\infty}^{+\infty} f^2(t) dt < \infty \right\}$ (i.e. the set of square integrable functions)

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

1.9.2 Connection of inner product to angle

In \mathbb{R}^n , the notion of inner product has a geometric interpretation, and is closely related to the notion of angle between vectors.

Definition:

$$\cos \theta = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} = \left\langle \frac{\mathbf{x}}{\|\mathbf{x}\|_2}, \frac{\mathbf{y}}{\|\mathbf{y}\|_2} \right\rangle \quad (2)$$

- $\langle \mathbf{x}, \mathbf{y} \rangle = 0 \Rightarrow \cos \theta = 0 \Rightarrow \theta = \frac{\pi}{2}$ (i.e. perpendicular)
- $\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \Rightarrow \cos \theta = 1 \Rightarrow \theta = 0$ (i.e. \mathbf{x} and \mathbf{y} are aligned)
- $\langle \mathbf{x}, \mathbf{y} \rangle = -\|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \Rightarrow \cos \theta = -1 \Rightarrow \theta = \pi$ (i.e. \mathbf{x} and \mathbf{y} are in opposite directions)
- $\langle \mathbf{x}, \mathbf{y} \rangle > 0 \Rightarrow \cos \theta > 0 \Rightarrow$ angle is acute
- $\langle \mathbf{x}, \mathbf{y} \rangle < 0 \Rightarrow \cos \theta < 0 \Rightarrow$ angle is obtuse

Derivation: L3: Inner products and orthogonality.

1.9.3 Cauchy-Schwartz inequality and its generalization

Definition:

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \quad (3)$$

Hölder's Inequality (generalization):

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q \quad \text{where } 1 \leq p, q < \infty \text{ and } \frac{1}{p} + \frac{1}{q} = 1 \quad (4)$$

Example: For $p = 1$ and $q = \infty$, we have:

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\|_1 \cdot \|\mathbf{y}\|_\infty$$

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \left(\sum_{i=1}^n |x_i| \right) \cdot \max_i |y_i|$$

1.9.4 Inner product induces a norm

Definition: Any inner product induces a norm, but not all norms are induced by an inner product.

- **Key:** If given an inner product, take the square root of the inner product to get the norm.
 - e.g. $\|\mathbf{x}\|_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$, which holds for \mathbb{R}^n and \mathbb{C}^n



Figure 3: Ordering of the vector spaces.

Warning: A norm doesn't induce an inner product (e.g. l_1 or l_∞)

1.10 Orthogonal decomposition

1.10.1 Mutually orthogonal

Definition: A set of non-zero vectors $S = \{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(d)}\}$ is **mutually orthogonal** if $\langle \mathbf{v}^{(i)}, \mathbf{v}^{(j)} \rangle = 0 \forall i \neq j$.

- **Fact:** Orthogonal set of vectors $S = \{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(d)}\}$ is linearly independent.
 - **Proof:** In L3.

1.10.2 Orthonormal basis

Definition: Set of orthogonal basis vectors that have unit norm.

If $S = \{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(d)}\}$ is a set of mutually orthogonal vectors, then $\left\{ \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \dots, \frac{\mathbf{v}_d}{\|\mathbf{v}_d\|} \right\}$ is an orthonormal basis for $\text{span}(S)$

Example: Standard basis is an orthonormal basis for \mathbb{R}^n

1.10.3 Orthogonal

Definition: Consider $\mathbf{x} \in \mathcal{V}$, and let S be a subspace of \mathcal{V} . We say \mathbf{x} is orthogonal to S if:

$$\langle \mathbf{x}, \mathbf{v} \rangle = 0 \quad \forall \mathbf{v} \in S.$$

We write: $\mathbf{x} \perp S$.

1.10.4 Orthogonal complement

Definition: The **orthogonal complement** of S , denoted S^\perp , is the set of all orthogonal vectors to S :

$$S^\perp = \{\mathbf{x} \in \mathcal{V} : \mathbf{x} \perp S\}$$

- S^\perp is a subspace. (Closed under addition and scalar multiplication)
- $S \cap S^\perp = \{\mathbf{0}\}$
- **Orthogonal decomposition:** Any $\mathbf{x} \in \mathcal{V}$ can be uniquely written as: $\mathbf{x} = \mathbf{x}_S + \mathbf{x}_{S^\perp}$ where $\mathbf{x}_S \in S$ and $\mathbf{x}_{S^\perp} \in S^\perp$

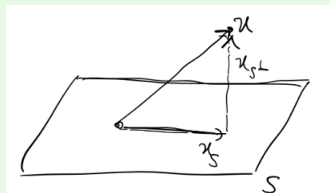


Figure 4: Drawing any \mathbf{x} .

- $\mathcal{V} = S + S^\perp = \{\mathbf{u} + \mathbf{v} : \mathbf{u} \in S, \mathbf{v} \in S^\perp\}$

2 Orthogonal Decomposition, Projecting onto Subspaces, Gram-Schmidt, QR Decomposition, Hyperplanes and Half-Spaces (Ch. 2.2-2.3)

2.1 Projection onto subspaces

Definition:

$$x^* = \text{Proj}_S(x) = \arg \min_{y \in S} \|x - y\|_2 \quad (5)$$

If $\{v^{(1)}, \dots, v^{(d)}\}$ is an orthonormal basis of S then

$$x^* = \sum_{i=1}^d \langle x, v^{(i)} \rangle v^{(i)} \quad (6)$$

- The error vector should be orthogonal to each vector in the subspace.

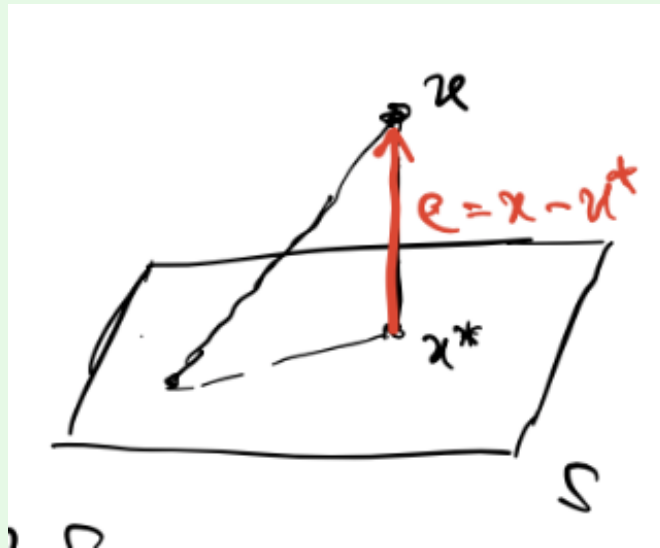


Figure 5: Error vector being perp. to S.

Example: For $v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$

The i th component can be extracted by doing the inner product with the i th standard basis:

$$\begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = v_1$$

$$\begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = v_2$$

$$\begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = v_3$$

Therefore, analogous to x^* , we can write them as the sum of the inner product times the standard basis.

$$v = v_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + v_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

2.1.1 Basic problem

Intuition: Given $x \in \mathcal{V}$ and a subspace S . Find the closest point (in norm) in S to x :

$$\text{Proj}_S(x) = \arg \min_{y \in S} \|y - x\| \quad (7)$$

- $\|y - x\|$: Some norm.
- **Subspace:** S doesn't have to be a subspace.
- **arg min:** Vector y that minimizes $\|x - y\|$

2.1.2 Projection onto a 1D subspace

Example: Projection onto a 1-dimensional subspace.

Let $S = \text{span}(\mathbf{v})$, and we denote the projection of \mathbf{x} onto S as:

$$\text{Proj}_S(\mathbf{x}) = \mathbf{x}^*$$

Under the Euclidean norm (i.e. ℓ_2 norm), we have nice geometry: we should have

$$\langle \mathbf{x} - \mathbf{x}^*, \mathbf{v} \rangle = 0$$

Since $\mathbf{x}^* \in S$, $\mathbf{x}^* = \alpha \mathbf{v}$ for some scalar α .

We need to find α .

So,

$$\begin{aligned} \langle \mathbf{x} - \alpha \mathbf{v}, \mathbf{v} \rangle &= 0 \\ \Rightarrow \langle \mathbf{x}, \mathbf{v} \rangle - \alpha \langle \mathbf{v}, \mathbf{v} \rangle &= 0 \\ \Rightarrow \alpha &= \frac{\langle \mathbf{x}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} \end{aligned}$$

Thus,

$$\mathbf{x}^* = \alpha \mathbf{v} = \frac{\langle \mathbf{x}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} \mathbf{v}$$

which simplifies to:

$$\mathbf{x}^* = \frac{\mathbf{x}^\top \mathbf{v}}{\|\mathbf{v}\|_2^2} \mathbf{v} = \left\langle \mathbf{x}, \frac{\mathbf{v}}{\|\mathbf{v}\|_2} \right\rangle \frac{\mathbf{v}}{\|\mathbf{v}\|_2}$$

- **Orthonormal Basis for S:** $\left\{ \frac{\mathbf{v}}{\|\mathbf{v}\|_2} \right\}$ since $\left\| \frac{\mathbf{v}}{\|\mathbf{v}\|_2} \right\|_2 = 1$
- **Projection Coefficient:** $\left\langle \mathbf{x}, \frac{\mathbf{v}}{\|\mathbf{v}\|_2} \right\rangle$
- **Note:** \mathbf{x}^* is the point we are looking for in the projection problem.

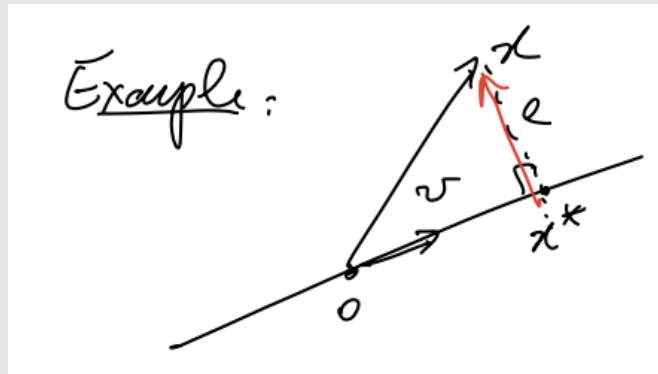


Figure 6: Visual representation of the projection problem.

2.1.3 Projection onto an n dimensional space

Example: This can be generalized to higher dimensions. Let S be a subspace of \mathcal{V} , and let $\{\mathbf{v}_1, \dots, \mathbf{v}_d\}$ be an orthonormal basis of S .

1. **Problem setup** Let

$$\mathbf{x}^* = \sum_{i=1}^d \alpha_i \mathbf{v}_i$$

Goal: Find $\alpha_1, \dots, \alpha_d$ so as to minimize the norm $\|\mathbf{x} - \mathbf{x}^*\|_2$.

2. **Derivation:** By geometry, we require that

$$\langle \mathbf{e}, \mathbf{v}_j \rangle = 0 \quad \forall j = 1, \dots, d$$

which implies:

$$\begin{aligned} \langle \mathbf{x} - \mathbf{x}^*, \mathbf{v}_j \rangle &= 0 \quad \forall j \\ \Rightarrow \langle \mathbf{x} - \sum_{i=1}^d \alpha_i \mathbf{v}_i, \mathbf{v}_j \rangle &= 0 \quad \forall j \end{aligned}$$

Using linearity of the inner product:

$$\Rightarrow \langle \mathbf{x}, \mathbf{v}_j \rangle = \sum_{i=1}^d \alpha_i \langle \mathbf{v}_i, \mathbf{v}_j \rangle$$

Since $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ if $i \neq j$ and 1 if $i = j$, this simplifies to:

$$\alpha_j = \langle \mathbf{x}, \mathbf{v}_j \rangle \quad \text{b/c only the } i=j \text{ term survives}$$

Thus,

$$\mathbf{x}^* = \sum_{i=1}^d \alpha_i \mathbf{v}_i = \sum_{i=1}^d \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{v}_i$$

3. **Solution:**

$$= \sum_{i=1}^d (\mathbf{x}^\top \mathbf{v}_i) \mathbf{v}_i$$

- $\mathbf{v}_i \in \mathbb{R}^n$.
- **Projection Coefficients:** $\mathbf{x}^\top \mathbf{v}_i$

4. **Example of Orthogonal Decomposition:**

$$\mathbf{e} = \mathbf{x} - \mathbf{x}^* \in S^\perp, \quad \mathbf{x}^* \in S$$

So,

$$\mathbf{x} = \mathbf{x}^* + \mathbf{e}, \quad \text{where } \mathbf{x}^* \in S, \quad \mathbf{e} \in S^\perp$$

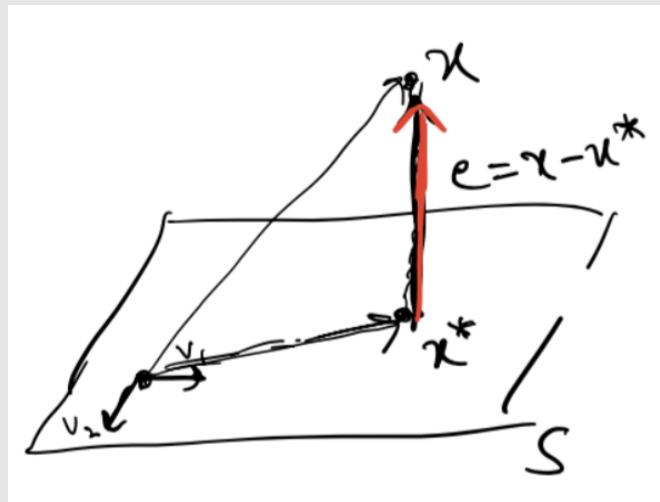


Figure 7: Generalization of projection.

2.1.4 Application of projections: Fourier series

Example: Fourier series:

1. Suppose we have a periodic function $x(t)$ with period T_0 .

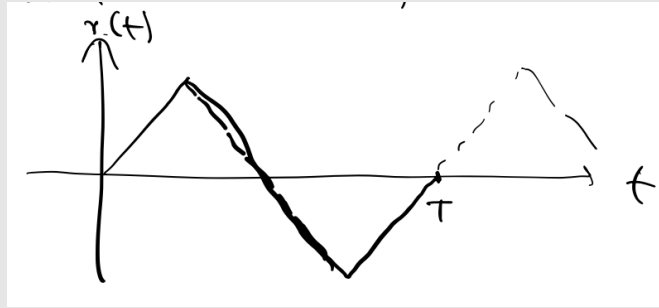


Figure 8: Periodic triangle function.

2. **Inner product for time domain (complex version):** $a_k = \langle x(t), y(t) \rangle = \frac{1}{T} \int_T x(t) \overline{y(t)} dt$

- **Note:** Real version is without the conjugate.

3. **Projection (i.e. one component of the sum):** $\text{Proj}_{\underline{v}_i}(\underline{x}) = \langle \underline{x}, \underline{v}_i \rangle \underline{v}_i$

4. **Goal:** Express $x(t)$ (i.e. any periodic function) as a sum of complex exponentials:

$$x^*(t) = \sum_{k=-\infty}^{\infty} a_k e^{jk\omega_0 t}$$

- **Projection:** $\text{Proj}_{e^{jk\omega_0 t}}(x(t)) = \langle x(t), \exp(jk\omega_0 t) \rangle e^{jk\omega_0 t} = a_k e^{jk\omega_0 t}$ for a certain value of k .

- **Projection coefficient:** $a_k = \langle x(t), e^{jk\omega_0 t} \rangle = \frac{1}{T_0} \int_0^T x(t) e^{-jk\omega_0 t} dt$

- **Fundamental frequency:** $\omega_0 = \frac{2\pi}{T_0}$.

5. **Prove orthonormal basis for the complex exponentials:** To prove it's a orthonormal basis, must prove it has unit norm 1 and each pair of vectors are orthogonal (i.e. inner product is 0).

- (a) **Magnitude of exp:** $|e^{j\theta}| = 1$. Therefore, it has unit norm.

- (b) **Orthogonality:**

$$\langle e^{ji\omega_0 t}, e^{jl\omega_0 t} \rangle = \begin{cases} 1, & i = l \\ 0, & i \neq l \end{cases}$$

Therefore, for each pair of basis vectors, they are orthogonal.

- **Conjugate of exp:** $(e^{j\theta})^* = e^{-j\theta}$

6. **Conclusion:** Fourier series is a projection of a function onto the set of orthonormal basis functions $\exp(jk\omega_0 t)$, where k is an integer.

- **Optimal:** This projection is optimal as it minimizes the approximation error $\|x(t) - x^*(t)\|$, i.e.

$$\frac{1}{T} \int_0^T (x(t) - x^*(t))^2 dt$$

As the number of terms in the summation increases to infinity, the error goes to 0.

2.2 Gram-Schmidt and QR decomposition

2.2.1 What if the set of basis vectors is not orthonormal?

Intuition: Let $\{u^{(1)}, \dots, u^{(d)}\}$ be a set of basis vectors for a subspace S (not necessarily orthonormal)

We can still use the orthogonality principle, i.e.,

$$e = x - x^* \perp S$$

Therefore,

$$\langle x - x^*, u^{(j)} \rangle = 0 \quad \forall j = 1, \dots, d$$

Also, $x^* \in S$ so x^* can be written as a linear combination of basis vectors, so $x^* = \sum_{i=1}^d \alpha_i u^{(i)}$

Need to find $\alpha_1, \dots, \alpha_d$ s.t.

$$\langle x - \sum_{i=1}^d \alpha_i u^{(i)}, u^{(j)} \rangle = 0 \quad \forall j = 1, \dots, d$$

$$\Rightarrow \langle x, u^{(j)} \rangle = \sum_{i=1}^d \alpha_i \langle u^{(i)}, u^{(j)} \rangle \quad \forall j = 1, \dots, d$$

$$\begin{bmatrix} \langle u^{(1)}, u^{(1)} \rangle & \langle u^{(2)}, u^{(1)} \rangle & \dots & \langle u^{(d)}, u^{(1)} \rangle \\ \langle u^{(1)}, u^{(2)} \rangle & \langle u^{(2)}, u^{(2)} \rangle & \dots & \langle u^{(d)}, u^{(2)} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle u^{(1)}, u^{(d)} \rangle & \langle u^{(2)}, u^{(d)} \rangle & \dots & \langle u^{(d)}, u^{(d)} \rangle \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_d \end{bmatrix} = \begin{bmatrix} \langle x, u^{(1)} \rangle \\ \langle x, u^{(2)} \rangle \\ \vdots \\ \langle x, u^{(d)} \rangle \end{bmatrix}$$

Solve for $\alpha_1, \dots, \alpha_d$, Then, we get $x^* = \sum_{i=1}^d \alpha_i u^{(i)}$

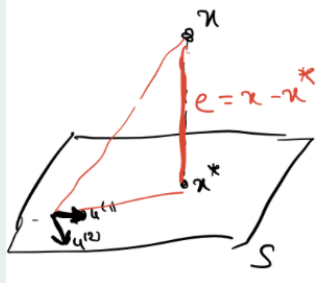


Figure 9: Not orthogonal, but similar to projection with orthonormal basis.

- **Note:** If $\{u^{(1)}, \dots, u^{(d)}\}$ is an orthonormal basis, then the matrix is the identity matrix, and we get $\alpha_j = \langle x, u^{(j)} \rangle$ as before.

Example: Function approximation. Let B be the set of basis functions that is not orthonormal:

$$\mathcal{B} = \{1, t, \dots, t^d\}$$

Let $x(t)$ be a function over $[0, 1]$.

- **1st Goal** Approximate $x(t)$ by $x^*(t) = \sum_{n=0}^d \alpha_n t^n$
- To find $\alpha_0, \alpha_1, \dots, \alpha_d$, need to solve the $Ax = b$.
- **2nd Goal:** Minimize the approximation error $\|x(t) - x^*(t)\|_2 = \left(\int_0^1 (x(t) - x^*(t))^2 dt \right)^{1/2}$

Recall: Taylor series expansion

$$x(t) \approx x(0) + x'(0)t + \frac{x''(0)}{2}t^2 + \dots$$

- Taylor series expansion is completely different from the projection method, and the reason is that Taylor series expansion is a local approximation.

2.2.2 Gram-Schmidt Procedure

Motivation: This is to get an orthonormal basis, so we can use the easier projection method.

Intuition: Another way to find the projection of x onto $S = \text{span}\{u^{(1)}, \dots, u^{(d)}\}$ is to first find an orthonormal basis of S , and then the projection problem becomes easier.

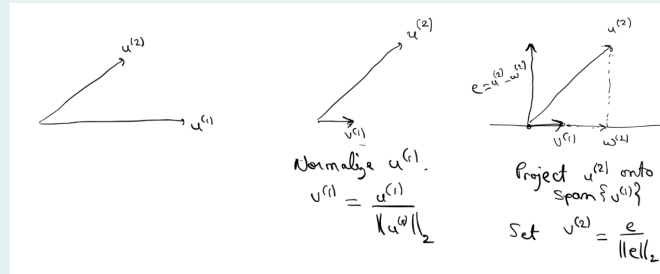


Figure 10: Gram-Schmidt Process for 2D.

1. Normalize $u^{(1)}$
2. Find the error vector by projecting $u^{(2)}$ onto the subspace $v^{(1)}$.
3. Normalize the error vector.
4. Now you have two vectors that form an orthonormal basis in 2D.

- 2.2.3 QR decomposition
- 2.2.4 Eigen-composition
- 2.2.5 Singular value decomposition
- 2.3 Hyperplanes and half-spaces
- 2.4 Non-euclidean projection
- 3 Non-Euclidean Projection, Projection onto Affine Sets, Functions, Gradients and Hessians (Ch. 2.3-2.4)
 - 3.1 Projection onto affine sets
 - 3.2 Functions
 - 3.3 Gradients
 - 3.4 Hessians
- 4 Matrices, Range, Null Space, Eigenvalues, Eigenvectors, Matrix Diagonalization (Ch. 3.1-3.5)
 - 4.1 Matrices
 - 4.2 Range
 - 4.3 Null Space
 - 4.4 Eigenvalues and eigenvectors
 - 4.5 Matrices diagonalization
- 5 Symmetric Matrices, Orthogonal Matrices, Spectral Decomposition, Positive Semidefinite Matrices, Ellipsoids (Ch. 4.1-4.4)
 - 5.1 Symmetric matrices
 - 5.2 Orthogonal matrices
 - 5.3 Spectral decomposition
 - 5.4 Positive semidefinite matrices
 - 5.5 Ellipsoids
- 6 Singular Value Decomposition, Principal Component Analysis (Ch. 5.1, 5.3.2)
 - 6.1 Singular value decomposition
 - 6.2 Principle component analysis
- 7 Interpretations of SVD, Low-Rank Approximation (Ch. 5.2-5.3.1)
 - 7.1 Interpretation of SVD
 - 7.2 Low-rank approximation
- 8 Least Squares, Overdetermined and Underdetermined Linear Equations (Ch. 6.1-6.4)
 - 8.1 Least squares
 - 8.2 Overdetermined linear equations