MA 110
Linear Algebra and Differential Equations
Lecture 20

Prof. Sudhir R. Ghorpade
Department of Mathematics
IIT Bombay
http://www.math.iitb.ac.in/∼srg/

Spring 2025

Least Squares Approximation

Suppose a large number of data points $(a_1, b_1), \ldots, (a_n, b_n)$ in $\mathbb{R}^2$ are given, and we want to find a straight line passing through these points. If all these points are collinear, then we may join any two of them by a straight line, which will work for us. If, however, these points are not collinear (which is most often the case), then we want to find a straight line $t = x_1 + s\,x_2$ in the st-plane which is most suitable in the following sense.

Consider the 'error' $|x_1 + a_j x_2 - b_j|$ caused because of the straight line $t = x_1 + s\,x_2$ not passing through the data point (a_j, b_j) for $j = 1, \ldots, n$. We collect these errors and attempt to find $x_1, x_2 \in \mathbb{R}$ such that the **least squares error**

$$\left(\sum_{j=1}^{n} |x_1 + a_j x_2 - b_j|^2 \right)^{1/2}.$$

is minimized. This is known as the **least squares problem**.

To solve this problem, let

$$\mathbf{A} = \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \vdots \\ 1 & a_n \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}. \text{ Then } \mathbf{Ax} = \begin{bmatrix} x_1 + a_1 x_2 \\ x_1 + a_2 x_2 \\ \vdots \\ x_1 + a_n x_2 \end{bmatrix}.$$

The least squares problem is to find $\mathbf{x} \in \mathbb{R}^{2\times 1}$ such that

$$\|\mathbf{Ax} - \mathbf{b}\|^2 = \sum_{j=1}^{n} |x_1 + a_j x_2 - b_j|^2$$

is minimised, that is, to find the best approximation to the vector $\mathbf{b}$ from the column space $\mathcal{C}(\mathbf{A}) = \{\mathbf{Ax} : \mathbf{x} \in \mathbb{R}^{2\times 1}\}$ of $\mathbf{A}$. To solve this, methods discussed earlier can be applied. An example is given in the tutorial problem 7.11 (i).

Abstract Vector Spaces

Throughout the course so far, we have dealt with row vectors, column vectors and matrices. We have introduced many interesting concepts such as linear independence of vectors, span of a set of vectors, a subspace of vectors, a basis for a subspace, the dimension of a subspace, the nullity and the rank of a matrix, linear transformations induced by matrices, inner product of two vectors, orthogonality of vectors, orthonormal basis for a subspace, orthogonal projection onto a subspace, and so on.

Based on these concepts, we have proved some important theorems like the Rank-Nullity theorem, the Fundamental Theorem for Linear Systems, the Spectral Theorem, the Projection Theorem.

Now we discuss these concepts in a more abstract setting.

Abstract Vector Space

Definition Let $\mathbb{K}$ denote $\mathbb{R}$ or $\mathbb{C}$ as usual. A **vector space** over $\mathbb{K}$ is a nonempty set V together with the operation of **addition** (i.e., a map $V \times V \to V$ given by $(u, v) \mapsto u + v$) and of **scalar multiplication** (i.e., a map $\mathbb{K} \times V \to V$ given by $(\alpha, v) \mapsto \alpha v$) satisfying the following properties.

I Closure axioms

1. $u + v \in V$ for all $u, v \in V$.
2. $\alpha \cdot v \in V$ for all $\alpha \in \mathbb{K}$ and $v \in V$.

(We shall write αv instead of $\alpha \cdot v \in V$ hence onward.)

II Axioms for addition

1. $u + v = v + u$ for all $u, v \in V$. (commutativity)
2. $u+(v+w)=(u+v)+w$ for all $u, v, w \in V$. (associativity)
3. There is (unique) $0 \in V$ such that $v + 0 = v$ for all $v \in V$.
4. For $v \in V$, there is (unique) $u \in V$ such that $v + u = 0$.

(We shall write this element u as $-v$.)

III Axioms for scalar multiplication

1. $\alpha(\beta v) = (\alpha\beta)v$ for all $\alpha, \beta \in \mathbb{K}$ and $v \in V$.
2. $\alpha(u + v) = \alpha u + \alpha v$ for all $\alpha \in \mathbb{K}$ and $u, v \in V$.
3. $(\alpha + \beta)v = \alpha v + \beta v$ for all $\alpha, \beta \in \mathbb{K}$ and $v \in V$.
4. $1v = v$ for all $v \in V$.

An element of a vector space is called a **vector**.

Examples: 1 $\mathbb{K}^n = \mathbb{K}^{n\times 1}$ is a vector space over $\mathbb{K}$. More generally, every vector subspace of $\mathbb{K}^n$ is a vector space over $\mathbb{K}$. Likewise for $\mathbb{K}^{1\times n}$.

2 $\mathbb{K}^{m\times n}$, the space of all $m \times n$ matrices with entries in $\mathbb{K}$, is a vector space over $\mathbb{K}$ with respect to the addition and scalar multiplication of matrices.

3 The set $\mathbb{K}[x]$ of all polynomials in the indeterminate x with coefficients in $\mathbb{K}$ is a vector space over $\mathbb{K}$ with respect to the usual addition and scalar multiplication of polynomials.

4 The set $\mathcal{P}_n$ of polynomials in $\mathbb{K}[x]$ of degree $\leq n$ is a vector space with addition and scalar multiplication as in 3.

5 Let $a, b \in \mathbb{R}$ with $a < b$. Then the space $C[a, b]$ of all continuous functions from $[a, b]$ to $\mathbb{R}$ a vector space over $\mathbb{R}$ with respect to pointwise addition and pointwise scalar multiplication of functions.

6 Let $a, b \in \mathbb{R}$ with $a < b$. Then the space $C^1[a, b]$ of all continuously differentiable functions from $[a, b]$ to $\mathbb{R}$ a vector space over $\mathbb{R}$ with addition and scalar multiplication as in 5 .

7 The space c of all convergent sequences of real numbers is a vector space over $\mathbb{R}$ with respect to pointwise addition and pointwise scalar multiplication of sequences.

Exercise: Verify that the above spaces are indeed vector spaces, i.e., all the axioms in the definition are satisfied.

Definition

Let V be a vector space (over $\mathbb{K}$). A nonempty subset W of V is called a **subspace** *of V if $v + w \in W$ for all $v, w \in W$ and $\alpha w \in W$ for all $\alpha \in \mathbb{K}$ and $w \in W$.*

Definition

Let V be a vector space (over $\mathbb{K}$), and let $n \in \mathbb{N}$. Given $v_1, \ldots, v_n \in V$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{K}$, the element

$$\alpha_1 v_1 + \cdots + \alpha_n v_n$$

of V is called the **linear combination** *of the vectors $v_1, \ldots, v_n$ with* **coefficients** *$\alpha_1, \ldots, \alpha_n$.*

Let V be a vector space. If W is a subspace of V, then clearly every linear combination of the elements of W belongs W.

Let $W_1. W_2$ be subspaces of V. Then it is easy to see that:

- $W_1 \cap W_2$ is a subspace of V. In fact it is the largest subspace of V which is contained in both W_1 and W_2.
- $W_1 \cup W_2$ need not be a subspace of V.
- $W_1 + W_2 := \{w_1 + w_2 : w_1 \in W_1 \text{ and } w_2 \in W_2\}$ is a subspace of V. In fact it is the smallest subspace of V containing both W_1 and W_2.

The notions in $\mathbb{R}^n$ of span of a set of vectors, linear dependence and independence of vectors, the dimension of a subspace of vectors carry over to an abstract vector space without any difficulty. Let V be a vector space (over $\mathbb{K}$).

Definition

Let $S \subset V$. The set of all linear combinations of elements of S is called the **span** *of S, and we denote it by* span S.

Definition

A subset S of V is called **linearly dependent** *if there are $v_1, \ldots, v_m$ in S and there are $\alpha_1, \ldots, \alpha_m \in \mathbb{K}$, not all zero, satisfying*

$$\alpha_1 v_1 + \cdots + \alpha_m v_m = 0.$$

It can be seen that S is linearly dependent $\iff$ either $\mathbf{0} \in S$ or a vector in S is a linear combination of other vectors in S.

Definition

A subset S of V is called **linearly independent** *if it is not linearly dependent, that is,*

$$\alpha_1 v_1 + \cdots + \alpha_m v_m = 0 \implies \alpha_1 = \cdots = \alpha_m = 0,$$

whenever $v_1, \ldots, v_m \in S$ and $\alpha_1, \ldots, \alpha_m \in \mathbb{K}$.

Clearly, S is linearly independent if and only if $\mathbf{0} \notin S$ and no elements of S is a linear combination of other elements of S.

The following crucial result was proved for the case $V := \mathbb{R}^{n\times 1}$. Exactly the same proof works in the general case.

Proposition

Let S be a subset of s elements and R be a set of r elements of V. If $S \subset \operatorname{span} R$ and $s > r$, then S is linearly dependent.

Examples

1. Let $m, n \in \mathbb{N}$, and let $V := \mathbb{K}^{m \times n}$ be set of all $m \times n$ matrices with entries in $\mathbb{K}$ with entry-wise addition and scalar multiplication. For $j = 1, \ldots, m$ and $k = 1, \ldots, n$, let $\mathbf{E}_{jk}$ denote the $m \times n$ matrix whose (j, k)th entry is equal to 1 and all other entries are equal to zero. Then the set $S := \{\mathbf{E}_{jk} : 1 \leq j \leq m, \, 1 \leq k \leq, n\}$ is linearly independent.

2. Let $V := c_0$ denote the set of all sequences in $\mathbb{K}$ which converge to 0. For $j \in \mathbb{N}$, let e_j denote the element of S whose j-th term is equal to 1 and all other terms are equal to 0. Then the set $S := \{e_j : j \in \mathbb{N}\}$ is linearly independent. Next, let $S_1 := S \cup \{e\}$, where the nth entry of the sequence e is equal to $1/n$ for $n \in \mathbb{N}$. Then the set S_1 is also linearly independent since e is not a (finite) linear combination of elements of S.

3. Let $V := \mathbb{K}[x]$ denote the set of all polynomials in the indeterminate x with coefficients in $\mathbb{K}$. Then the set $S := \{x^j : j = 0, 1, 2, \ldots\}$ is linearly independent.
Next, let $S_1 := S \cup \{p\}$, where $p \in \mathbb{K}[x]$. Then the set S_1 is linearly dependent since $p \in \operatorname{span} S$.

4. Let $V := C[-\pi, \pi]$. For $n \in \mathbb{N}$, let

$$u_n(t) := \cos nt \quad \text{and} \quad v_n(t) := \sin nt \quad \text{for } t \in [-\pi, \pi].$$

Then the set $S := \{u_1, u_2, \ldots\} \cup \{v_1, v_2, \ldots\}$ is linearly independent. (Note that the zero element of this vector space is the function having all its values on $[-\pi, \pi]$ equal to 0.) To prove this, use the idea that if $\alpha \cos t + \beta \sin t = 0$, then by differentiating, we also have $-\alpha \sin t + \beta \cos t = 0$.

Next, let $S_1 := S \cup \{w\}$, where $w(t) := t$ for $t \in [-\pi, \pi]$. Then the set S_1 is also linearly independent, since $w(\pi) \neq w(-\pi)$, and so $w \notin \operatorname{span} S$.

Definition

A vector space V is said to be **finite dimensional** *if there is a finite subset S of V such that $V = \text{span}\, S$; otherwise the vector space V is said to be* **infinite dimensional**.

If a vector space V is infinite dimensional, then V is larger than the span of any finite subset of V, and so V must contain an infinite linearly independent subset. Conversely, if V contains an infinite linearly independent subset, then V must be infinite dimensional.

Examples: Let $n, m \in \mathbb{N}$. The vector spaces $\mathbb{K}^{n\times 1}$, $\mathbb{K}^{1\times n}$ and $\mathbb{K}^{m\times n}$ are finite dimensional, and so is the vector space $\mathcal{P}_n$ of all polynomials in the indeterminate x having degree less than or equal to n. But the vector spaces $\mathbb{K}[x]$, $C[-\pi, \pi]$, c, and c_0 are infinite dimensional.

Definition

Any linearly independent subset of a finite dimensional vector space V which spans V is called a **basis** *for V.*

Here is the most important result about finite dimensional vector spaces. The proof is similar to that in the case of subspaces of $\mathbb{K}^n$.

Proposition

Let V be a finite dimensional vector space over $\mathbb{K}$. Then the following holds.

- *V has a basis.*
- *Every set that spans V has a subset which is a basis of V.*
- *Every linearly independent subset of V can be extended to a basis of V.*
- *Any two bases of V have the same cardinality, called the* **dimension** *of V and denoted by* $\dim V$.