01204211 Discrete Mathematics
Lecture 12b: Linear functions

Jittat Fakcharoenphol

September 15, 2022

Linear functions

Linear functions

Consider vector spaces $\mathcal{V}$ and $\mathcal{W}$ over $\mathbb{R}$. A function $f : \mathcal{V} \to \mathcal{W}$ is **linear** if

1. for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}$, $f(\boldsymbol{x} + \boldsymbol{y}) = f(\boldsymbol{x}) + f(\boldsymbol{y})$ and
2. for all $\alpha \in \mathbb{R}$ and $\boldsymbol{x} \in \mathcal{V}$, $f(\alpha \boldsymbol{x}) = \alpha f(\boldsymbol{x})$.

Example 1 - MLP

Example 2 - Page rank (1)

Example 2 - Page rank (2)

Matrix-vector multiplication

Given an $m \times n$ matrix M over $\mathbb{R}$, consider a product

$$M\boldsymbol{x}.$$

Note that for the multiplication to work, $\boldsymbol{x}$ must be in $\mathbb{R}^n$ and the result vector is in $\mathbb{R}^m$. Therefore, we can define a function $f : \mathbb{R}^n \to \mathbb{R}^m$ as

$$f(\boldsymbol{x}) = M\boldsymbol{x}.$$

Note that f is linear because:

$$f(\boldsymbol{x}+\boldsymbol{y}) = M(\boldsymbol{x}+\boldsymbol{y}) = M\boldsymbol{x} + M\boldsymbol{y} = f(\boldsymbol{x}) + f(\boldsymbol{y}),$$

and

$$f(\alpha\boldsymbol{x}) = M(\alpha\boldsymbol{x}) = \alpha M\boldsymbol{x} = \alpha f(\boldsymbol{x}).$$

The converse

Lemma 1

For any linear function $f : \mathbb{R}^n \to \mathbb{R}^m$*, there exists an* $m \times n$ *matrix* M *such that*

$$f(\boldsymbol{x}) = M\boldsymbol{x}.$$

Proof.

Consider any $x \in \mathbb{R}^n$. Let $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]$. Note that

$$\boldsymbol{x} = [x_1, 0, \ldots, 0] + [0, x_2, 0, \ldots, 0] + \cdots + [0, \ldots, 0, x_n].$$

Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n \in \mathbb{R}^n$ be standard generators, i.e., $\boldsymbol{e}_i$ be a vector with 1 at the i-th row and 0 at every other positions. (For example $\boldsymbol{e}_1 = [1, 0, \ldots, 0]$ and $\boldsymbol{e}_3 = [0, 0, 1, 0, \ldots, 0]$.)
We thus have

$$\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_n\boldsymbol{e}_n.$$

Since f is linear, this implies that

$$f(\boldsymbol{x}) = x_1 f(\boldsymbol{e}_1) + x_2 f(\boldsymbol{e}_2) + \cdots + x_n f(\boldsymbol{e}_n).$$

□

Proof (cont.)

Define M as follows

$$M = \left[\begin{array}{c|c|c|c} f(\boldsymbol{e}_1) & f(\boldsymbol{e}_2) & \cdots & f(\boldsymbol{e}_n) \end{array}\right].$$

Hence,

$$\begin{aligned} M\boldsymbol{x} &= \left[\begin{array}{c|c|c|c} f(\boldsymbol{e}_1) & f(\boldsymbol{e}_2) & \cdots & f(\boldsymbol{e}_n) \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= x_1 f(\boldsymbol{e}_1) + x_2 f(\boldsymbol{e}_2) + \cdots + x_n f(\boldsymbol{e}_n) = f(\boldsymbol{x}), \end{aligned}$$

as required. $\square$

Structures of linear functions

Zero

Lemma 2

Consider any linear function $f : \mathcal{V} \to \mathcal{W}$*. Let* $0_{\mathcal{V}}$ *denote the zero vector in* $\mathcal{V}$ *and* $0_{\mathcal{W}}$ *denote the zero vector in* $\mathcal{W}$*. We have that linear function* f *always maps zero to zero, i.e.,* $f(0_{\mathcal{V}}) = 0_{\mathcal{W}}$*.*

Proof.

First note that $0_{\mathcal{V}} = 0_{\mathcal{V}} + 0_{\mathcal{V}}$. Since f is linear, we have that

$$f(0_{\mathcal{V}}) = f(0_{\mathcal{V}} + 0_{\mathcal{V}}) = f(0_{\mathcal{V}}) + f(0_{\mathcal{V}}).$$

Subtracting $f(0_{\mathcal{V}})$ from both sides, we conclude that

$$0_{\mathcal{W}} = f(0_{\mathcal{V}}).$$

□

One-to-one linear functions and Onto linear functions

One-to-one and onto functions

Consider a function $f : D \to R$ (i.e., the domain of f is D and its range is R).

- Function f is **one-to-one** (or **injective**) if for all $x, y \in D$, $f(x) = f(y)$ implies that $x = y$.
- Function f is **onto** (or **surjective**) if for all $x \in R$, there exists $y \in D$ such that $f(y) = x$.

For this course, we consider only linear functions; therefore, we consider $f : \mathcal{V} \to \mathcal{W}$, where $\mathcal{V}$ and $\mathcal{W}$ are vector spaces.

One-to-one linear functions

Suppose that f is not one-to-one, i.e., there exists a pair $x, y \in \mathcal{V}$ such that $x \neq y$ and $f(x) = f(y)$. Since f is linear, we know that

$$f(x - y) = f(x) - f(y) = 0.$$

Since $x \neq y$, $x - y \neq 0$ and we have that there exists a non-zero element $z = x - y$ that f maps to 0. The contraposition of this fact is as follows.
If the only element in $\mathcal{V}$ that f maps to $0_{\mathcal{W}}$ is $0_{\mathcal{V}}$, f is one-to-one (or injective).

Because the set of elements that f maps to zero is very important, we have a name for it.

Definition (Kernel)

The **kernel** of f, denoted by Ker f, is the set of all elements that f maps to zero, i.e.,

$$\text{Ker } f = \{\boldsymbol{v} \in \mathcal{V} : f(\boldsymbol{v}) = 0_{\mathcal{V}}\}.$$

We can now restate the condition for f to be one-to-one using this concept.

Lemma 3

A linear function f is one-to-one, if and only if Ker $f = \{0\}$.

The kernel is also a vector space

Lemma 4

Ker f *is a vector space.*

Proof.

First note that $f(0) = 0$; thus $0 \in$ Ker f.
Suppose that $x \in$ Ker f, i.e., $f(x) = 0$. Note that for any scalar α,

$$f(\alpha x) = \alpha f(x) = \alpha 0 = 0.$$

Also suppose $y \in$ Ker f. We have that

$$f(x + y) = f(x) + f(y) = 0 + 0 = 0.$$

$\square$

Onto linear functions

Definition (Image)

For any function g, its **image**, denoted by $\operatorname{Im} g$, is the set of all elements that g maps to, i.e.,

$$\operatorname{Im} g = \{y : \text{there exists } x \text{ such that } g(x) = y\}.$$

The image is also a vector space

Lemma 5

The image of linear function f, $\mathrm{Im}\ f$, is a vector space.

Proof.

Since $f(0_{\mathcal{V}}) = 0_{\mathcal{W}}$, $0_{\mathcal{W}} \in \mathrm{Im}\ f$.
Consider $y \in \mathrm{Im}\ f$. We have that there exists x such that $f(x) = y$. Consider any scalar α. We know that $\alpha y \in \mathrm{Im}\ f$ because $f(\alpha x) = \alpha f(x) = \alpha y$.
Consider, also, $y' \in \mathrm{Im}\ f$. Let x' be such that $f(x') = y'$. Since $y' \in \mathrm{Im}\ f$, we know that x' exists. We have that

$$f(x + x') = f(x) + f(x') = y + y'.$$

This implies that $y + y' \in \mathrm{Im}\ f$. □

Kernels and images

Theorem 6 (Kernel-Image Theorem)

Consider a linear function $f : \mathcal{V} \to \mathcal{W}$*. We have that*

$$\dim \mathcal{V} = \dim \mathsf{Ker}\ f + \dim \mathsf{Im}\ f.$$

Completing the basis

Lemma 7

For a set of linearly independent vectors

$$\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k$$

in $\mathcal{V}$ with basis $B = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ (where $k \leq n$), there exists a set of vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{n-k} \in B$ such that

$$\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k, \boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{n-k}\}$$

is also a basis for $\mathcal{V}$.

Proof.

Use the morphing lemma. □

Theorem 8 (Kernel-Image Theorem)

For a linear function $f : \mathcal{V} \to \mathcal{W}$, $\dim \mathcal{V} = \dim \mathsf{Ker}\ f + \dim \mathsf{Im}\ f$.

Proof of Kernel-Image Theorem (1).

Let $n = \dim \mathcal{V}$ and $k = \dim \mathsf{Ker}\ f$. Our goal is to show that $\dim \mathsf{Im}\ f = n - k$.

Since $\mathsf{Ker}\ f$ is a vector space, there is a basis $B = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$. From the previous slide, we can find other $n - k$ vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{n-k}$ to extend B to be a basis S for $\mathcal{V}$, i.e., we have that

$$S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k, \boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{n-k}\}$$

is a basis for $\mathcal{V}$. □

Proof of Kernel-Image Theorem (2).

Consider any $\boldsymbol{u} \in \mathcal{V}$. We can write $\boldsymbol{u}$ as

$$\alpha_1\boldsymbol{v}_1 + \alpha_2\boldsymbol{v}_2 + \cdots + \alpha_k\boldsymbol{v}_k + \beta_1\boldsymbol{w}_1 + \beta_2\boldsymbol{w}_2 + \cdots + \beta_{n-k}\boldsymbol{w}_{n-k},$$

because S is a basis. Consider $f(\boldsymbol{u})$. We have that

$$\begin{aligned} f(\boldsymbol{u}) &= f(\alpha_1\boldsymbol{v}_1 + \cdots + \alpha_k\boldsymbol{v}_k + \beta_1\boldsymbol{w}_1 + \cdots + \beta_{n-k}\boldsymbol{w}_{n-k}) \\ &= f(\alpha_1\boldsymbol{v}_1) + \cdots + f(\alpha_k\boldsymbol{v}_k) + f(\beta_1\boldsymbol{w}_1) + \cdots + f(\beta_{n-k}\boldsymbol{w}_{n-k}) \\ &= f(\beta_1\boldsymbol{w}_1) + f(\beta_2\boldsymbol{w}_2) + \cdots + f(\beta_{n-k}\boldsymbol{w}_{n-k}) \\ &= \beta_1 f(\boldsymbol{w}_1) + \beta_2 f(\boldsymbol{w}_2) + \cdots + \beta_{n-k} f(\boldsymbol{w}_{n-k}) \end{aligned}$$

(Note that the second step follows because $\boldsymbol{v}_i \in \text{Ker } f$. Other steps use the fact that f is linear.)

This calculation shows that an image of f can be written as a linear combination of $f(\boldsymbol{w}_1), \ldots, f(\boldsymbol{w}_{n-k})$. That is

$$\text{Im } f = \operatorname{Span}\{f(\boldsymbol{w}_1), \ldots, f(\boldsymbol{w}_{n-k})\}.$$

$\square$

Proof of Kernel-Image Theorem (3).

Let $S' = \{f(\boldsymbol{w}_1), \ldots, f(\boldsymbol{w}_{n-k})\}$. If we can show that S' is a basis for Im f, we are done because that would imply that $\dim \text{Im } f = n - k$ as required.

We already know that S' spans Im f. To show that S' is a basis we still need to show that S' is linearly independent.

Suppose that there exist $\beta_1, \ldots, \beta_{n-k}$ such that

$$\beta_1 f(\boldsymbol{w}_1) + \beta_2 f(\boldsymbol{w}_2) + \cdots + \beta_{n-k} f(\boldsymbol{w}_{n-k}) = 0_{\mathcal{W}}.$$

Since f is linear we know that

$$\begin{aligned} 0_{\mathcal{W}} &= \beta_1 f(\boldsymbol{w}_1) + \beta_2 f(\boldsymbol{w}_2) + \cdots + \beta_{n-k} f(\boldsymbol{w}_{n-k}) \\ &= f(\beta_1 \boldsymbol{w}_1) + f(\beta_2 \boldsymbol{w}_2) + \cdots + f(\beta_{n-k} \boldsymbol{w}_{n-k}) \\ &= f(\beta_1 \boldsymbol{w}_1 + \beta_2 \boldsymbol{w}_2 + \cdots + \beta_{n-k} \boldsymbol{w}_{n-k}), \end{aligned}$$

i.e., $\beta_1 \boldsymbol{w}_1 + \beta_2 \boldsymbol{w}_2 + \cdots + \beta_{n-k} \boldsymbol{w}_{n-k}$ is in Ker f. $\square$

Proof of Kernel-Image Theorem (4).

Suppose that some $\beta_i \neq 0$.
Since

$$\beta_1 \boldsymbol{w}_1 + \beta_2 \boldsymbol{w}_2 + \cdots + \beta_{n-k} \boldsymbol{w}_{n-k} \in \mathsf{Ker}\ f,$$

we know that it is a linear combination of vectors from B, as B is a basis for vector space Ker f.
From here, we can reach a contradiction using the fact that vectors in S are linearly independent.

Therefore, we conclude that all $\beta_1, \ldots, \beta_{n-k}$ must be 0. Hence, $S' = \{f(\boldsymbol{w}_1), \ldots, f(\boldsymbol{w}_{n-k})\}$ is linearly independent as needed. $\square$

Ranks and nullities

Direct sum (optional)

Consider two subspaces $\mathcal{V}$ and $\mathcal{W}$ of a vector space $\mathcal{Z}$. If $\mathcal{V} \cap \mathcal{W} = \{0\}$, we can define their *direct sum* to be another vector space $\mathcal{V} \oplus \mathcal{W}$ as

$$\mathcal{V} \oplus \mathcal{W} = \{\boldsymbol{v} + \boldsymbol{u} : \boldsymbol{v} \in \mathcal{V}, \boldsymbol{u} \in \mathcal{W}\}.$$

Note, again, that $\mathcal{V} \oplus \mathcal{W}$ is defined only when $\mathcal{V} \cap \mathcal{W} = \{0\}$.