Theory sheet 4

Linear forms

Definition 1. *Let* $a_1, \dots, a_n \in \mathbb{R}$ *be numbers. A* <u>*linear form*</u> *associated with* $(a_1, \dots, a_n)$ *is a function* $f : \mathbb{R}^n \to \mathbb{R}$ *given by*

$$f(x_1, \dots, x_n) = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = \sum_{k=1}^{n} a_k x_k.$$

If we denote $A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ *and* $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ *, when* f *may be written as*

$$f(\boldsymbol{x}) = A^\top \boldsymbol{x}.$$

For example,

$$f(x_1, x_2) = \begin{pmatrix} 3 \\ 4 \end{pmatrix}^\top \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 3x_1 + 4x_2.$$

Remark 1. *Column vectors* $=$ $n \times 1$ *matrices are usually written either as* $\boldsymbol{x}$ *(in bold) or as* $\vec{x}$ *(with arrow on top). Both notations are standard. The second one is more frequently used in handwriting, whereas the second is more standard in print.*

Here is an alternative definition:

Definition 2. *A linear form* $f : \mathbb{R}^n \to \mathbb{R}$ *is a function satisfying*

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

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ *.*

The two definitions are <u>equivalent</u>:

Theorem 1. *If* $f(\boldsymbol{x} + \boldsymbol{y}) = f(\boldsymbol{x}) + f(\boldsymbol{y})$ *for all* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ *, there exist numbers* $a_1, \dots, a_n$ *such that* $f(\boldsymbol{x}) = a_1 x_1 + \cdots + a_n x_n$ *.*

The column vector A^T has a special name:

Definition 3. *If* $f : \mathbb{R}^n \to \mathbb{R}$ *is a linear form* $f(\boldsymbol{x}) = A^\top \boldsymbol{x}$ *, then* $A^\top$ *is called the* <u>*gradient*</u> *of* f *.*

Later on, we shall introduce gradients of other functions.

Definition 4. *A graph of a function* $f : \mathbb{R}^n \to \mathbb{R}$ *is the set*

$$\left\{(x_1, \ldots, x_n, y) \in \mathbb{R}^{n+1} : y = f(x_1, \ldots, x_n)\right\}.$$

Definition 5 (Line, plane, hyperplane)**.**

- *A line passing through the origin is the graph of a linear form* $f : \mathbb{R} \to \mathbb{R}$.
- *A plane passing through the origin is the graph of a linear form* $f : \mathbb{R}^2 \to \mathbb{R}$.
- *A hyperplane passing through the origin is the graph of a linear form* $f : \mathbb{R}^n \to \mathbb{R}$.

Check that this corresponds to your intuitive understanding of what line and plane are by drawing a few pictures.

Definition 6. *If* $f(\boldsymbol{x}) = a_1x_1 + a_2x_2 + \cdots + a_nx_n$, *then* a_i *is said to be the slope of* f *in the direction* x_i.

Level curves/sets of linear forms

Definition 7. *The level set of a function* $f : \mathbb{R}^n \to \mathbb{R}$ *with value* $c \in \mathbb{R}$ *is a set where this function takes value* c*:*

$$\{(x_1, \ldots, x_n) : f(x_1, \ldots, x_n) = c\}.$$

Two remarks:

- Level sets of linear forms are lines, planes or hyperplanes not necessarily passing through zero.
- The gradient is orthogonal to the level set.

Linear transformation

Definition 8. *Let* $A \in M_{m,n}$ *be a matrix. A linear transformation associated with* A *is a function* $f : \mathbb{R}^n \to \mathbb{R}^m$ *given by*

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

More explicitly,

$$f(x_1, \ldots, x_n) = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix},$$

where numbers y_i *are given by*

$$y_i = (A\boldsymbol{x})_i = \sum_{k=1}^{n} a_{ik}x_k.$$

As with linear forms, there is an alternative definition:

Definition 9. *A linear transformation is a function* $f : \mathbb{R}^n \to \mathbb{R}^m$ *which satisfies*

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

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ *.*

And as with linear forms, this definition is equivalent to the first one:

Theorem 2. *If* $f(\boldsymbol{x}+\boldsymbol{y}) = f(\boldsymbol{x}) + f(\boldsymbol{y})$ *for all* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ *, then there exists a matrix* $A \in M_{m,n}$ *such that* $f(\boldsymbol{x}) = A\boldsymbol{x}$ *.*

Proof. Since $f(\boldsymbol{x}+\boldsymbol{y}) = f(\boldsymbol{x}) + f(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$f(\boldsymbol{x}) = f\left(\sum_{j=1}^{n} x_j \boldsymbol{e}_j\right) = \sum_{j=1}^{n} x_j f(\boldsymbol{e}_j),$$

where $\boldsymbol{e}_j$ is the j^{th} vector of the standard basis:

$$\boldsymbol{e}_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \leftarrow i$$

Therefore, the i^{th} component of $\boldsymbol{y} = f(\boldsymbol{x})$ is given by

$$y_i = \sum_{j=1}^{n} x_j (f(\boldsymbol{e}_j))_i = \sum_{j=1}^{n} A_{ij} x_j,$$

where $A_{ij} = (f(\boldsymbol{e}_j))_i$. □

Here are a few remarks:

- Note the position of m and n : if $A \in M_{m,n}$, then $f(\boldsymbol{x}) = A\boldsymbol{x}$ is a function $\mathbb{R}^n \to \mathbb{R}^m$.
- In other words, m is the output dimension and n is the input dimension.
- The mysterious definition of matrix product actually comes from the following fact: if $f : \mathbb{R}^n \to \mathbb{R}^p$ and $g : \mathbb{R}^p \to \mathbb{R}^m$ satisfy

  $$f(\boldsymbol{x}+\boldsymbol{y}) = f(\boldsymbol{x}) + f(\boldsymbol{y}) \quad \text{and} \quad g(\boldsymbol{x}'+\boldsymbol{y}') = g(\boldsymbol{x}') + g(\boldsymbol{y}')$$

  for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and all $\boldsymbol{x}', \boldsymbol{y}' \in \mathbb{R}^p$, then their composition

  $$h : \mathbb{R}^n \to \mathbb{R}^m : h(\boldsymbol{x}) = g(f(\boldsymbol{x}))$$

  also satisfies this property:

  $$h(\boldsymbol{x}+\boldsymbol{y}) = h(\boldsymbol{x}) + h(\boldsymbol{y})$$

  for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. We know from the above theorem that $h(\boldsymbol{x}) = C\boldsymbol{x}$ with some C , but is it possible to find C ?

- If f is represented by matrix B (that is, $f(\boldsymbol{x}) = B\boldsymbol{x}$) and g is represented by matrix A (that is, $g(\boldsymbol{x}') = A\boldsymbol{x}'$), then h is represented by $C = AB$.
- In other words, product of matrices represents composition of linear transformations.
- Compare:

$$A \in M_{m,p},\ B \in M_{p,n} \implies AB \in M_{m,n}$$

$$f : \mathbb{R}^n \to \mathbb{R}^p,\ g : \mathbb{R}^p \to \mathbb{R}^m \implies g(f(\boldsymbol{x})) : \mathbb{R}^n \to \mathbb{R}^m.$$

Examples of linear transformations

Rotations

If $\theta \in [0, 2\pi]$ is some angle, the linear transformation $r_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ associated with the matrix R_θ by $r_\theta(\boldsymbol{x}) = R_\theta \boldsymbol{x}$, where

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

rotates $\boldsymbol{x}$ by the angle θ about $\mathbf{0}$.

For example, if $\theta = -\frac{\pi}{3}$, we have

$$R_{-\pi/3} = \begin{pmatrix} \cos\left(-\frac{\pi}{3}\right) & -\sin\left(-\frac{\pi}{3}\right) \\ \sin\left(-\frac{\pi}{3}\right) & \cos\left(-\frac{\pi}{3}\right) \end{pmatrix} = \begin{pmatrix} 1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & 1/2 \end{pmatrix}.$$

Dilation/homothethy/scaling

If $k \in \mathbb{R}$, the linear transformation $h : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$h(\boldsymbol{x}) = \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

is the dilation or homothethy or scaling by a factor of k.

- If $|k| > 1$, h increases sizes
- If $|k| < 1$, h decreases sizes
- If $k > 0$, h preserves directions
- If $k < 0$, h flips directions

Solving linear systems

Given $\boldsymbol{y} \in \mathbb{R}^m$ and $A \in M_{m,n}$, we want to solve $A\boldsymbol{x} = \boldsymbol{y}$ for $\boldsymbol{x}$. There are two natural questions:

- Is this possible? In other words, does there exist $\boldsymbol{x}$ such that $A\boldsymbol{x} = \boldsymbol{y}$?
- Is $\boldsymbol{x}$ unique? In other words, is it possible that $A\boldsymbol{x} = \boldsymbol{y} = A\boldsymbol{x}'$ with $\boldsymbol{x} \neq \boldsymbol{x}'$?

Remark 2. *In the previous lecture we have seen that $A\boldsymbol{x} = \boldsymbol{y}$ may be solved by $\boldsymbol{x} = A^{-1}\boldsymbol{y}$, but only if A^{-1} exists. Recall that A^{-1} only makes sense for square matrices. Now we are interested in solving $A\boldsymbol{x} = \boldsymbol{y}$ without assuming that the matrix is square.*

The following theorem gives the answer (without proof):

Theorem 3.

- *If rank of A coincides with the output dimension, that is, $\operatorname{Rank} A = m$, then $\boldsymbol{x}$ exists*
- *If rank of A is smaller than the output dimension, that is, $\operatorname{Rank} A < m$, then **there exist** some $\boldsymbol{y} \in \mathbb{R}^m$ such that equation $A\boldsymbol{x} = \boldsymbol{y}$ is unsolvable*
- *If rank of A coincides with the input dimension, that is, $\operatorname{Rank} A = n$, then the solution is unique (if exists; see previous points)*
- *If rank of A is smaller than the input dimension, then the solution is not unique (if exists; see previous points)*

In particular,

> If rank of A coincides with both input and output dimensions
>
> $$\operatorname{Rank} A = m = n,$$
>
> the solution exists and is unique.

Here are a few remarks:

- There are three possibilities either (1) there is a unique solution or (2) are no solutions or (3) there is infinite number of solutions.
- Example of case (1):

$$\begin{cases} x_1 + x_2 = 2 \\ x_1 - x_2 = 0 \end{cases} \implies x_1 = x_2 = 1.$$

- Example of case (2):

$$\begin{cases} x_1 + x_2 = 1 \\ x_1 + x_2 = 2 \end{cases} \implies \text{no solutions}$$

- Example of case (3):

$$\begin{cases} x_1 + x_2 = 1 \\ 2x_1 + 2x_2 = 2 \end{cases} \implies x_2 = 1 - x_1, \quad x_2 \text{ any number.}$$

- If $\operatorname{Rank} A < m$ (rank smaller than the output dimension), $A\boldsymbol{x} = \boldsymbol{y}$ may be solvable for some $\boldsymbol{y}$, just not all $\boldsymbol{y}$.
- If $\operatorname{Rank} A < n$ (rank smaller than the inpu dimension), there may be many solutions and we usually want to describe the entire family of solutions.

Solving linear systems using elementary transformations

In order to solve a linear system

$$\begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & y_1 \\ a_{21}x_2 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & y_2 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_m & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & y_m \end{cases}$$

we rewrite this system of equations in the matrix form

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

and use the same idea we have used for finding rank, determinant and inverse: apply elementary transformations to both sides of this equation (meaning to A and to $\boldsymbol{y}$).

The goal is the same as with finding of the inverse: to transform the matrix on the left into the identity matrix I.

However, this may not be possible if A is not invertible. If $\operatorname{Rank} A < m$, at some point we encounter zero rows:

$$i \rightarrow \begin{pmatrix} * & * & * & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & * \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} * \\ \vdots \\ \tilde{y}_i \\ \vdots \\ * \end{pmatrix} \leftarrow i$$

- If the i^{th} value $\tilde{y}_i$ is not zero, no choice of $\boldsymbol{x}$ will produce a solution. In this case we conclude that there is no solution.
- If $\tilde{y}_i = 0$, we can just throw away the zero line and proceed with solving the problem.

If, on the other hand, $\operatorname{Rank} A < n$, at some point we may encounter zero columns:

$$\begin{pmatrix} * & \dots & 0 & \dots & * \\ \vdots & \vdots & \vdots & \dots & \vdots \\ * & \dots & 0 & \dots & * \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} * \\ \vdots \\ * \end{pmatrix}.$$
$$\uparrow j$$

- If the j^{th} column is zero, every choice of $x_j \in \mathbb{R}$ gives a solution.
- Therefore, the solution (if exists) is non-unique.

Remark 3. *In practice, we frequently do not need to simplify the matrix to the end. Let us take an example. If we have reduced the system to the form*

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \\ 2 \end{pmatrix},$$

we immediately see that the problem does not have solutions because there is non-zero value 2 *against a zero row. If, on the other hand,*

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \\ 0 \end{pmatrix},$$

then we can, without simplifying the matrix further, get back to the "system of equations" form

$$\begin{cases} x_1 + x_2 + x_3 = 2 \\ x_2 + 3x_3 = 6 \end{cases}$$

and solve it by hand:

$$x_2 = 6 - 3x_3, \quad x_1 = 2 - x_3 - (6 - 3x_3) = -4 + 2x_3,$$

where x_3 *is any number.*

As we did with finding the inverse, it is useful to introduce the augmented matrix notation for solving linear systems:

Definition 10. *Let us associate with a system of linear equations*

$$\begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & y_1 \\ a_{21}x_2 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & y_2 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_m & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & y_m \end{cases}$$

an augmented matrix $(A \mid \boldsymbol{y})$ *written as*

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & y_1 \\ a_{21} & a_{22} & \dots & a_{2n} & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & y_m \end{array}\right).$$

Similarly to what we did to find the inverse, we can now solve this system of linear equations by applying elementary transformations to the augmented matrix $(A \mid \boldsymbol{y})$.