# ECON2125/4021/8013

## Lecture 8

John Stachurski

Semester 1, 2015

# Linear Maps

In this section we investigate one of the most important classes of functions

These are the so-called linear functions

Linear functions play a fundamental role in all fields of science

- In one-to-one correspondence with matrices

Even nonlinear functions can often be rewritten as partially linear

The properties of linear functions are closely tied to notions such as

- linear combinations, span
- linear independence, bases, etc.

# Linearity

A function $T\colon \mathbb{R}^K \to \mathbb{R}^N$ is called **linear** if

$$T(\alpha \mathbf{x} + \beta \mathbf{y}) = \alpha T\mathbf{x} + \beta T\mathbf{y} \qquad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^K,\ \forall\, \alpha, \beta \in \mathbb{R}$$

Notation:

- Linear functions often written with upper case letters
- Typically omit parenthesis around arguments when convenient

Example. $T \colon \mathbb{R} \to \mathbb{R}$ defined by $Tx = 2x$ is linear

Proof: Take any $\alpha, \beta, x, y$ in $\mathbb{R}$ and observe that

$$T(\alpha x + \beta y) = 2(\alpha x + \beta y) = \alpha 2x + \beta 2y = \alpha Tx + \beta Ty$$

Example. The function $f \colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is <u>non</u>linear

Proof: Set $\alpha = \beta = x = y = 1$

Then

- $f(\alpha x + \beta y) = f(2) = 4$
- But $\alpha f(x) + \beta f(y) = 1 + 1 = 2$

## Example

Given constants $c_1$ and $c_2$, the function $T\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

$$T\mathbf{x} = T(x_1, x_2) = c_1 x_1 + c_2 x_2$$

is linear

Proof: If we take any $\alpha, \beta$ in $\mathbb{R}$ and $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^2$, then

$$\begin{aligned} T(\alpha\mathbf{x} + \beta\mathbf{y}) &= c_1[\alpha x_1 + \beta y_1] + c_2[\alpha x_2 + \beta y_2] \\ &= \alpha[c_1 x_1 + c_2 x_2] + \beta[c_1 y_1 + c_2 y_2] \\ &= \alpha T\mathbf{x} + \beta T\mathbf{y} \end{aligned}$$

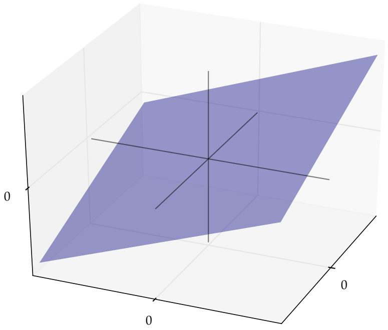



Figure : The graph of $T\mathbf{x} = c_1 x_1 + c_2 x_2$ is a plane through the origin

Remark: Thinking of linear functions as those whose graph is a straight line is not correct

## Example

Function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 1 + 2x$ is <u>non</u>linear

Proof: Take $\alpha = \beta = x = y = 1$

Then

- $f(\alpha x + \beta y) = f(2) = 5$
- But $\alpha f(x) + \beta f(y) = 3 + 3 = 6$

This kind of function is called an **affine** function

Let $\mathbf{a}_1, \ldots, \mathbf{a}_K$ be vectors in $\mathbb{R}^N$

Let $T \colon \mathbb{R}^K \to \mathbb{R}^N$ be defined by

$$T\mathbf{x} = T \begin{pmatrix} x_1 \\ \vdots \\ x_K \end{pmatrix} = x_1 \mathbf{a}_1 + \ldots + x_K \mathbf{a}_K$$

**Ex.** Show that this function is linear

Remarks

- This is a generalization of the previous linear examples
- In a sense it is the most general representation of a linear map from $\mathbb{R}^K$ to $\mathbb{R}^N$
- It is also "the same" as the $N \times K$ matrix with columns $\mathbf{a}_1, \ldots, \mathbf{a}_K$ — more on this later

# Implications of Linearity

**Fact.** If $T \colon \mathbb{R}^K \to \mathbb{R}^N$ is a linear map and $\mathbf{x}_1, \ldots, \mathbf{x}_J$ are vectors in $\mathbb{R}^K$, then for any linear combination we have

$$T\left[\alpha_1 \mathbf{x}_1 + \cdots + \alpha_J \mathbf{x}_J\right] = \alpha_1 T\mathbf{x}_1 + \cdots + \alpha_J T\mathbf{x}_J$$

Proof for $J = 3$: Applying the def of linearity twice,

$$\begin{aligned} T\left[\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \alpha_3 \mathbf{x}_3\right] &= T\left[(\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2) + \alpha_3 \mathbf{x}_3\right] \\ &= T\left[\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2\right] + \alpha_3 T\mathbf{x}_3 \\ &= \alpha_1 T\mathbf{x}_1 + \alpha_2 T\mathbf{x}_2 + \alpha_3 T\mathbf{x}_3 \end{aligned}$$

**Ex.** Show that if $T$ is any linear function then $T\mathbf{0} = \mathbf{0}$

**Fact.** If $T\colon \mathbb{R}^K \to \mathbb{R}^N$ is a linear map, then

$$\operatorname{rng}(T) = \operatorname{span}(V) \quad \text{where} \quad V := \{T\mathbf{e}_1, \ldots, T\mathbf{e}_K\}$$

- Here $\mathbf{e}_k$ is the $k$-th canonical basis vector in $\mathbb{R}^K$

Proof: Any $\mathbf{x} \in \mathbb{R}^K$ can be expressed as $\sum_{k=1}^K \alpha_k \mathbf{e}_k$

Hence $\operatorname{rng}(T)$ is the set of all points of the form

$$T\mathbf{x} = T\left[\sum_{k=1}^K \alpha_k \mathbf{e}_k\right] = \sum_{k=1}^K \alpha_k T\mathbf{e}_k$$

as we vary $\alpha_1, \ldots, \alpha_K$ over all combinations

This coincides with the definition of $\operatorname{span}(V)$

## Example

Let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$T\mathbf{x} = T(x_1, x_2) = x_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ -2 \end{pmatrix}$$

Then

$$T\mathbf{e}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \text{and} \quad T\mathbf{e}_2 = \begin{pmatrix} 0 \\ -2 \end{pmatrix}$$

**Ex.** Show that $V := \{T\mathbf{e}_1, T\mathbf{e}_2\}$ is linearly independent

We conclude that the range of $T$ is all of $\mathbb{R}^2$ (why?)

The **null space** or **kernel** of linear map $T \colon \mathbb{R}^K \to \mathbb{R}^N$ is

$$\ker(T) := \{\mathbf{x} \in \mathbb{R}^K : T\mathbf{x} = \mathbf{0}\}$$

**Ex.** Show that $\ker(T)$ is a linear subspace of $\mathbb{R}^K$

**Fact.** $\ker(T) = \{\mathbf{0}\}$ if and only if $T$ is one-to-one

Proof of $\implies$ : Suppose that $T\mathbf{x} = T\mathbf{y}$ for arbitrary $\mathbf{x}, \mathbf{y} \in \mathbb{R}^K$

Then $\mathbf{0} = T\mathbf{x} - T\mathbf{y} = T(\mathbf{x} - \mathbf{y})$

In other words, $\mathbf{x} - \mathbf{y} \in \ker(T)$

Hence $\ker(T) = \{\mathbf{0}\} \implies \mathbf{x} = \mathbf{y}$

# Linearity and Bijections

Many scientific and practical problems are “inverse” problems

- We observe outcomes but not what caused them
- How can we work backwards from outcomes to causes?

Examples

- What consumer preferences generated observed market behavior?
- What kinds of expectations led to given shift in exchange rates?

Loosely, we can express an inverse problem as

outcome

model

$$F(x) = y$$

what $x$ led to outcome $y$?

- Does this problem have a solution?
- Is it unique?

Answers depend on whether $F$ is one-to-one, onto, etc.

The best case is a bijection

But other situations also arise

Recall that an arbitrary function can be

- one-to-one
- onto
- both (a bijection)
- neither

For linear functions from $\mathbb{R}^N$ to $\mathbb{R}^N$, the first three are all equivalent!

In particular,

$$\text{onto} \iff \text{one-to-one} \iff \text{bijection}$$

The next theorem summarizes

**Fact.** If $T$ is a linear function from $\mathbb{R}^N$ to $\mathbb{R}^N$ then all of the following are equivalent:

1. $T$ is a bijection
2. $T$ is onto
3. $T$ is one-to-one
4. $\ker(T) = \{\mathbf{0}\}$
5. The set of vectors $V := \{T\mathbf{e}_1, \ldots, T\mathbf{e}_N\}$ is linearly independent

If any one of these equivalent conditions is true, then $T$ is called **nonsingular**

- Don't forget: We are talking about $\mathbb{R}^N$ to $\mathbb{R}^N$ here

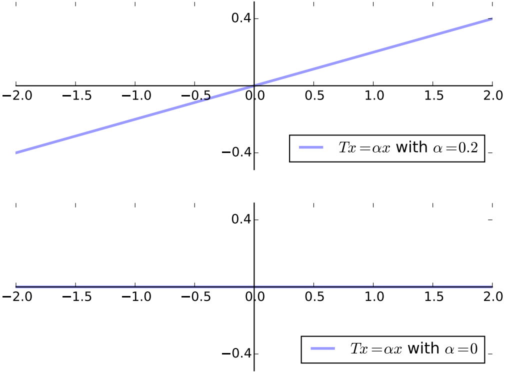



Figure : The case of $N = 1$, nonsingular and singular

Proof that $T$ onto $\iff V := \{T\mathbf{e}_1, \ldots, T\mathbf{e}_N\}$ is linearly independent

Recall that for any linear map $T$ we have $\mathrm{rng}(T) = \mathrm{span}(V)$

Using this fact and the definitions,

$$\begin{aligned} T \text{ onto } &\iff \mathrm{rng}(T) = \mathbb{R}^N \\ &\iff \mathrm{span}(V) = \mathbb{R}^N \\ &\iff V \text{ is linearly indepenent} \end{aligned}$$

(We saw that $N$ vectors span $\mathbb{R}^N$ iff linearly indepenent)

Rest of proof: Solved exercises

**Fact.** If $T \colon \mathbb{R}^N \to \mathbb{R}^N$ is nonsingular then so is $T^{-1}$.

What is the implication here?

If $T$ is a bijection then so is $T^{-1}$

Hence the only real claim is that $T^{-1}$ is also linear

The proof is an exercise...

# Maps Across Different Dimensions

Remember that these results apply to maps from $\mathbb{R}^N$ to $\mathbb{R}^N$

Things change when we look at linear maps across dimensions

The general rules for linear maps are

- Maps from lower to higher dimensions cannot be onto
- Maps from higher to lower dimensions cannot be one-to-one

In either case they cannot be bijections

The next fact summarizes

**Fact.** For a linear map $T$ from $\mathbb{R}^K \to \mathbb{R}^N$, the following statements are true:

1. If $K < N$ then $T$ is not onto
2. If $K > N$ then $T$ is not one-to-one

Proof of part 1: Let $K < N$ and let $T \colon \mathbb{R}^K \to \mathbb{R}^N$ be linear

Letting $V := \{T\mathbf{e}_1, \ldots, T\mathbf{e}_K\}$, we have

$$\dim(\mathrm{rng}(T)) = \dim(\mathrm{span}(V)) \leq K < N$$

$$\therefore \quad \mathrm{rng}(T) \neq \mathbb{R}^N$$

Hence $T$ is not onto

Proof of part 2: $K > N \implies T$ is not one-to-one

Suppose to the contrary that $T$ is one-to-one

Let $\alpha_1, \ldots, \alpha_K$ be a collection of vectors such that

$$\alpha_1 T\mathbf{e}_1 + \cdots + \alpha_K T\mathbf{e}_K = \mathbf{0}$$

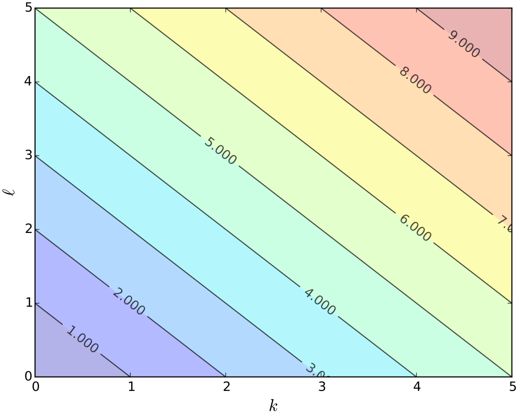
$$\therefore \quad T(\alpha_1 \mathbf{e}_1 + \cdots + \alpha_K \mathbf{e}_K) = \mathbf{0} \qquad \text{(by linearity)}$$

$$\therefore \quad \alpha_1 \mathbf{e}_1 + \cdots + \alpha_K \mathbf{e}_K = \mathbf{0} \qquad (\text{since } \ker(T) = \{\mathbf{0}\})$$

$$\therefore \quad \alpha_1 = \cdots = \alpha_K = 0 \qquad (\text{by independence of } \{\mathbf{e}_1, \ldots \mathbf{e}_K\})$$

We have shown that $\{T\mathbf{e}_1, \ldots, T\mathbf{e}_K\}$ is linearly independent

But then $\mathbb{R}^N$ contains a linearly independent set with $K > N$ vectors — contradiction




Example. Cost function $c(k, \ell) = rk + w\ell$ cannot be one-to-one

# Matrices and Linear Equations

We now begin our study of matrices

As we'll see, there's an isomorphic relationship between

1. matrices
2. linear maps

Often properties of matrices are best understood via those of linear maps

# Matrices

Typical $N \times K$ **matrix**:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NK} \end{pmatrix}$$

Symbol $a_{nk}$ stands for element in the

- $n$-th row
- $k$-th column

Often matrices correspond to coefficients of a linear equation

$$\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1K}x_K = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2K}x_K = b_2 \\ \vdots \\ a_{N1}x_1 + a_{N2}x_2 + \cdots + a_{NK}x_K = b_N \end{array} \tag{1}$$

Given the $a_{nm}$ and $b_n$, what values of $x_1, \ldots, x_K$ solve this system?

We now investigate this and other related questions

But first some background on matrices...

An $N \times K$ matrix also called a

- **row vector** if $N = 1$
- **column vector** if $K = 1$

Examples.

$$\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix} \text{ is } N \times 1, \qquad \mathbf{c} = \begin{pmatrix} c_1 \cdots c_K \end{pmatrix} \text{ is } 1 \times K$$

If $N = K$, then $\mathbf{A}$ is called **square**

We use

- $\mathrm{col}_k(\mathbf{A})$ to denote the $k$-th column of $\mathbf{A}$
- $\mathrm{row}_n(\mathbf{A})$ to denote the $n$-th row of $\mathbf{A}$

Example

$$\mathrm{col}_1(\mathbf{A}) = \mathrm{col}_1 \begin{pmatrix} a_{11} & \cdots & a_{1K} \\ a_{21} & \cdots & a_{2K} \\ \vdots & \vdots & \vdots \\ a_{N1} & \cdots & a_{NK} \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{N1} \end{pmatrix}$$

The **zero matrix** is

$$\mathbf{0} := \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

The **identity matrix** is

$$\mathbf{I} := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

# Algebraic Operations for Matrices

Addition and scalar multiplication are also defined for matrices

Both are element by element, as in the vector case

Scalar multiplication:

$$\gamma \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NK} \end{pmatrix} := \begin{pmatrix} \gamma a_{11} & \gamma a_{12} & \cdots & \gamma a_{1K} \\ \gamma a_{21} & \gamma a_{22} & \cdots & \gamma a_{2K} \\ \vdots & \vdots & & \vdots \\ \gamma a_{N1} & \gamma a_{N2} & \cdots & \gamma a_{NK} \end{pmatrix}$$

Addition:

$$
\begin{pmatrix}
a_{11} & \cdots & a_{1K} \\
a_{21} & \cdots & a_{2K} \\
\vdots & \vdots & \vdots \\
a_{N1} & \cdots & a_{NK}
\end{pmatrix}
+
\begin{pmatrix}
b_{11} & \cdots & b_{1K} \\
b_{21} & \cdots & b_{2K} \\
\vdots & \vdots & \vdots \\
b_{N1} & \cdots & b_{NK}
\end{pmatrix}
$$

$$
:= \begin{pmatrix}
a_{11} + b_{11} & \cdots & a_{1K} + b_{1K} \\
a_{21} + b_{21} & \cdots & a_{2K} + b_{2K} \\
\vdots & \vdots & \vdots \\
a_{N1} + b_{N1} & \cdots & a_{NK} + b_{NK}
\end{pmatrix}
$$

Note that matrices must be same dimension

Multiplication of matrices:

Product $\mathbf{AB}$: $i, j$-th element is inner product of $i$-th row of $\mathbf{A}$ and $j$-th column of $\mathbf{B}$

$$\begin{pmatrix} a_{11} & \cdots & a_{1K} \\ a_{21} & \cdots & a_{2K} \\ \vdots & \vdots & \vdots \\ a_{N1} & \cdots & a_{NK} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1J} \\ b_{21} & \cdots & b_{2J} \\ \vdots & \vdots & \vdots \\ b_{K1} & \cdots & b_{KJ} \end{pmatrix} = \begin{pmatrix} c_{11} & \cdots & c_{1J} \\ c_{21} & \cdots & c_{2J} \\ \vdots & \vdots & \vdots \\ c_{N1} & \cdots & c_{NJ} \end{pmatrix}$$

In this display,

$$c_{11} = \sum_{k=1}^{K} a_{1k} b_{k1}$$

Suppose $\mathbf{A}$ is $N \times K$ and $\mathbf{B}$ is $J \times M$

- $\mathbf{AB}$ defined only if $K = J$
- Resulting matrix $\mathbf{AB}$ is $N \times M$

The rule to remember:

$$\text{product of } N \times K \text{ and } K \times M \text{ is } N \times M$$

Important: Multiplication is not commutative

In particular, it is not in general true that $\mathbf{AB} = \mathbf{BA}$

- In fact $\mathbf{BA}$ is not well-defined unless $N = M$ also holds

Useful observation:

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NK} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{pmatrix}$$

$$= x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{N1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{N2} \end{pmatrix} + \cdots + x_K \begin{pmatrix} a_{1K} \\ a_{2K} \\ \vdots \\ a_{NK} \end{pmatrix}$$

$$= \sum_{k=1}^{K} x_k \operatorname{col}_k(\mathbf{A})$$

Rules for multiplication:

**Fact.** Given scalar $\alpha$ and conformable $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, we have

1. $\mathbf{A}(\mathbf{B}\mathbf{C}) = (\mathbf{A}\mathbf{B})\mathbf{C}$
2. $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{A}\mathbf{B}+\mathbf{A}\mathbf{C}$
3. $(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{A}\mathbf{C}+\mathbf{B}\mathbf{C}$
4. $\mathbf{A}\alpha\mathbf{B} = \alpha\mathbf{A}\mathbf{B}$

(Here "conformable" means operation makes sense)

The $k$**-th power** of a square matrix $\mathbf{A}$ is

$$\mathbf{A}^k := \underbrace{\mathbf{A}\cdots\mathbf{A}}_{k \text{ terms}}$$

If it exists, the **square root** of $\mathbf{A}$ is written $\mathbf{A}^{1/2}$

Defined as the matrix $\mathbf{B}$ such that $\mathbf{B}^2$ is $\mathbf{A}$

More on these later...

In matrix multiplication, $\mathbf{I}$ is the multiplicative unit

That is, assuming conformability, we always have

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A}$$

**Ex.** Check it using the definition of matrix multiplication

Note: If $\mathbf{I}$ is $K \times K$, then

$$\mathrm{col}_k(\mathbf{I}) = \mathbf{e}_k = \text{ } k\text{-th canonical basis vector in } \mathbb{R}^K$$

```
In [1]: import numpy as np

In [2]: A = [[2, 4],
   ...:      [4, 2]]

In [3]: A = np.array(A)  # Convert A to array

In [4]: B = np.identity(2)

In [5]: B
Out[5]:
array([[ 1.,  0.],
       [ 0.,  1.]])
```

```
In [6]: A + B         # Matrix addition
Out[6]:
array([[ 3.,  4.],
       [ 4.,  3.]])

In [7]: np.dot(A, B)  # Matrix multiplication
Out[7]:
array([[ 2.,  4.],
       [ 4.,  2.]])
```

# Matrices as Maps

Any $N \times K$ matrix $\mathbf{A}$ can be thought of as a function $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$

- In $\mathbf{A}\mathbf{x}$ the $\mathbf{x}$ is understood to be a column vector

It turns out that every such map is linear

To see this fix $N \times K$ matrix $\mathbf{A}$ and let $T$ be defined by

$$T \colon \mathbb{R}^K \to \mathbb{R}^N, \qquad T\mathbf{x} = \mathbf{A}\mathbf{x}$$

Pick any $\mathbf{x}$, $\mathbf{y}$ in $\mathbb{R}^K$, and any scalars $\alpha$ and $\beta$

The rules of matrix arithmetic tell us that

$$T(\alpha\mathbf{x} + \beta\mathbf{y}) := \mathbf{A}(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha\mathbf{A}\mathbf{x} + \beta\mathbf{A}\mathbf{y} =: \alpha T\mathbf{x} + \beta T\mathbf{y}$$

So matrices make linear functions

How about examples of linear functions that don't involve matrices?

Actually there are none!

**Fact.** If $T \colon \mathbb{R}^K \to \mathbb{R}^N$ then

$$T \text{ is linear} \iff \exists\, N \times K \text{ matrix } \mathbf{A} \text{ s.t. } T\mathbf{x} = \mathbf{A}\mathbf{x},\ \forall\, \mathbf{x} \in \mathbb{R}^K$$

- On the last slide we showed the $\Longleftarrow$ part
- On the next slide we show the $\Longrightarrow$ part

Let $T \colon \mathbb{R}^K \to \mathbb{R}^N$ be linear

We aim to construct an $N \times K$ matrix $\mathbf{A}$ such that

$$T\mathbf{x} = \mathbf{A}\mathbf{x}, \qquad \forall\, \mathbf{x} \in \mathbb{R}^K$$

As usual, let $\mathbf{e}_k$ be the $k$-th canonical basis vector in $\mathbb{R}^K$

Define a matrix $\mathbf{A}$ by $\operatorname{col}_k(\mathbf{A}) = T\mathbf{e}_k$

Pick any $\mathbf{x} = (x_1, \ldots, x_K) \in \mathbb{R}^K$

By linearity we have

$$T\mathbf{x} = T\left[\sum_{k=1}^K x_k \mathbf{e}_k\right] = \sum_{k=1}^K x_k T\mathbf{e}_k = \sum_{k=1}^K x_k \operatorname{col}_k(\mathbf{A}) = \mathbf{A}\mathbf{x}$$

# Matrix Product as Composition

Let

- $\mathbf{A}$ be $N \times K$ and $\mathbf{B}$ be $K \times M$
- $T \colon \mathbb{R}^K \to \mathbb{R}^N$ be the linear map $T\mathbf{x} = \mathbf{A}\mathbf{x}$
- $U \colon \mathbb{R}^M \to \mathbb{R}^K$ be the linear map $U\mathbf{x} = \mathbf{B}\mathbf{x}$

The matrix product $\mathbf{AB}$ corresponds exactly to the composition of $T$ and $U$

Proof:

$$(T \circ U)(\mathbf{x}) = T(U\mathbf{x}) = T(\mathbf{B}\mathbf{x}) = \mathbf{A}\mathbf{B}\mathbf{x}$$

This helps us understand a few things

For example, let

- $\mathbf{A}$ be $N \times K$ and $\mathbf{B}$ be $J \times M$
- $T \colon \mathbb{R}^K \to \mathbb{R}^N$ be the linear map $T\mathbf{x} = \mathbf{A}\mathbf{x}$
- $U \colon \mathbb{R}^M \to \mathbb{R}^J$ be the linear map $U\mathbf{x} = \mathbf{B}\mathbf{x}$

Then $\mathbf{AB}$ is only defined when $K = J$

This is because $\mathbf{AB}$ corresponds to $T \circ U$

But for $T \circ U$ to be well defined we need $K = J$

Then $U$ maps $\mathbb{R}^M$ to $\mathbb{R}^K$ and $T$ maps $\mathbb{R}^K$ to $\mathbb{R}^N$

## Column Space

Let $\mathbf{A}$ be an $N \times K$ matrix

The **column space** of $\mathbf{A}$ is defined as the span of its columns

$$\operatorname{span}(\mathbf{A}) = \operatorname{span}\{\operatorname{col}_1(\mathbf{A}), \ldots, \operatorname{col}_K(\mathbf{A})\}$$

$$= \text{all vectors of the form } \sum_{k=1}^{K} x_k \operatorname{col}_k(\mathbf{A})$$

Equivalently,

$$\operatorname{span}(\mathbf{A}) := \{\mathbf{A}\mathbf{x} : \mathbf{x} \in \mathbb{R}^K\}$$

This is exactly the range of the associated linear map

$$T \colon \mathbb{R}^K \to \mathbb{R}^N \text{ defined by } T\mathbf{x} = \mathbf{A}\mathbf{x}$$

Example. If

$$\mathbf{A} = \begin{pmatrix} 1 & -5 \\ 2 & 3 \end{pmatrix}$$

then the span is all linear combinations

$$x_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + x_2 \begin{pmatrix} -5 \\ 3 \end{pmatrix} \quad \text{where} \quad (x_1, x_2) \in \mathbb{R}^2$$

These columns are linearly independent (shown earlier)

Hence the column space is all of $\mathbb{R}^2$ (why?)

**Ex.** Show that the column space of any $N \times K$ matrix is a linear subspace of $\mathbb{R}^N$

# Rank

Equivalent questions

- How large is the range of the linear map $T\mathbf{x} = \mathbf{A}\mathbf{x}$?
- How large is the column space of $\mathbf{A}$?

The obvious measure of size for a linear subspace is its dimension

The dimension of $\mathrm{span}(\mathbf{A})$ is known as the **rank** of $\mathbf{A}$

$$\mathrm{rank}(\mathbf{A}) := \dim(\mathrm{span}(\mathbf{A}))$$

Because $\mathrm{span}(\mathbf{A})$ is the span of $K$ vectors, we have

$$\mathrm{rank}(\mathbf{A}) = \dim(\mathrm{span}(\mathbf{A})) \leq K$$

$\mathbf{A}$ is said to have **full column rank** if

$$\operatorname{rank}(\mathbf{A}) = \text{ number of columns of } \mathbf{A}$$

**Fact.** For any matrix $\mathbf{A}$, the following statements are equivalent:

1. $\mathbf{A}$ is of full column rank
2. The columns of $\mathbf{A}$ are linearly independent
3. If $\mathbf{A}\mathbf{x} = \mathbf{0}$, then $\mathbf{x} = \mathbf{0}$

**Ex.** Check this, recalling that

$$\dim(\operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_K\}) = K \iff \{\mathbf{a}_1, \ldots, \mathbf{a}_K\} \text{ linearly indepenent}$$

```
In [1]: import numpy as np

In [2]: A = [[2.0, 1.0],
   ...:      [6.3, 3.0]]

In [3]: np.linalg.matrix_rank(A)
Out[3]: 2

In [4]: A = [[2.0, 1.0],  # Col 2 is half col 1
   ...:      [6.0, 3.0]]

In [5]: np.linalg.matrix_rank(A)
Out[5]: 1
```