

# Lecture Notes 1: Matrix Algebra

## Part A: Vectors and Matrices

Peter J. Hammond

My email is `p.j.hammond@warwick.ac.uk`  
or `hammond@stanford.edu`

A link to these lecture slides can be found at  
`https://web.stanford.edu/~hammond/pjhLects.html`

revised 2020 September 14th

# Outline

## Solving Two Equations in Two Unknowns

### First Example

## Vectors

Vectors and Inner Products

Addition, Subtraction, and Scalar Multiplication

Linear versus Affine Functions

Norms and Unit Vectors

Orthogonality

The Canonical Basis

Linear Independence and Dimension

## Matrices

Matrices and Their Transposes

Matrix Multiplication: Definition

## Example of Two Equations in Two Unknowns

It is easy to check that

$$\left. \begin{array}{l} x + y = 10 \\ x - y = 6 \end{array} \right\} \implies x = 8, y = 2$$

More generally, one can:

1. add the two equations, to eliminate  $y$ ;
2. subtract the second equation from the first, to eliminate  $x$ .

This leads to the following transformation

$$\left. \begin{array}{l} x + y = b_1 \\ x - y = b_2 \end{array} \right\} \implies \begin{cases} 2x = b_1 + b_2 \\ 2y = b_1 - b_2 \end{cases}$$

of the two equation system with general right-hand sides.

Obviously the solution is

$$x = \frac{1}{2}(b_1 + b_2), y = \frac{1}{2}(b_1 - b_2)$$

## Using Matrix Notation, I

Matrix notation allows the two equations

$$1x + 1y = b_1$$

$$1x - 1y = b_2$$

to be expressed as

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

or as  $\mathbf{Az} = \mathbf{b}$ , where

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \mathbf{z} = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

Here  $\mathbf{A}$ ,  $\mathbf{z}$ ,  $\mathbf{b}$  are respectively: (i) the **coefficient matrix**;  
(ii) the **vector of unknowns**; (iii) the **vector of right-hand sides**.

## Using Matrix Notation, II

Also, the solution  $x = \frac{1}{2}(b_1 + b_2)$ ,  $y = \frac{1}{2}(b_1 - b_2)$  can be expressed as

$$x = \frac{1}{2}b_1 + \frac{1}{2}b_2$$

$$y = \frac{1}{2}b_1 - \frac{1}{2}b_2$$

or as

$$\mathbf{z} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \mathbf{C}\mathbf{b}, \quad \text{where } \mathbf{C} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

## Two General Equations

Consider the general system of two equations

$$\begin{aligned}ax + by &= u = 1u + 0v \\cx + dy &= v = 0u + 1v\end{aligned}$$

in two unknowns  $x$  and  $y$ , filled in with some extra 1s and 0s.

In matrix form, these equations can be written as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

For simplicity, we assume throughout the subsequent analysis that the coefficients  $a, b, c, d$  on the left-hand sides are all  $\neq 0$ .

## Three Different Cases

We are considering the two equations

$$ax + by = u \quad \text{and} \quad cx + dy = v$$

They correspond to the two straight lines

$$y = (u - ax)/b \quad \text{and} \quad y = (v - cx)/d$$

The two lines have respective slopes  $-a/b$  and  $-c/d$ .

These two slopes are equal iff  $a/b = c/d$ , or iff  $ad = bc$ , or iff  $D := ad - bc = 0$ .

We will distinguish three cases:

- (A) If  $D \neq 0$ , the two lines have different slopes, so their intersection consists of a single point.
- (B) If  $D = 0$  and  $u/b \neq v/d$ , then the two lines are parallel but distinct, so their intersection is empty.
- (C) If  $D = 0$  and  $u/b = v/d$ , then the two lines are identical, so their intersection consists of all the points on either line.

## First Steps

We assume that  $a \neq 0$  in the matrix equation

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

We can eliminate  $x$  from the second equation by adding  $-c/a$  times the first row to the second.

Given  $D = ad - bc$ , we obtain the new equality

$$\begin{pmatrix} a & b \\ 0 & D/a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -c/a & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

Now multiply the second row by  $a$  to obtain

$$\begin{pmatrix} a & b \\ 0 & D \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -c & a \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$



## Two General Equations: Case A

The equations are

$$\begin{pmatrix} a & b \\ 0 & D \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -c & a \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

In **Case A** when  $D := ad - bc \neq 0$ ,

we can add  $-b/D$  times the second row to the first, which yields

$$\begin{pmatrix} a & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 + (bc/D) & -ab/D \\ -c & a \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

Recognizing that  $1 + (bc/D) = (D + bc)/D = ad/D$ ,

then dividing the two rows/equations by  $a$  and  $D$  respectively, we obtain

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{D} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

This implies the unique solution

$$x = (1/D)(du - bv) \quad \text{and} \quad y = (1/D)(av - cu)$$

## Two General Equations: Cases B and C

Beyond Case A, when  $D := ad - bc = 0$ ,  
the multiplier  $-ab/D$  is undefined and the system

$$\begin{pmatrix} a & b \\ 0 & D/a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -c/a & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

collapses to  $\begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u \\ v - cu/a \end{pmatrix}$ .

This leaves us with two cases:

Case B) If  $cu \neq av$ , there is no solution.

We are trying to find the intersection of two distinct parallel lines.

Case C) If  $cu = av$ , then the second equation reduces to  $0 = 0$ .

There is a continuum of solutions

satisfying the one remaining equation  $ax + by = u$ ,

or  $x = (u - by)/a$  where  $y$  is any real number.

# Outline

## Solving Two Equations in Two Unknowns

First Example

## Vectors

### Vectors and Inner Products

Addition, Subtraction, and Scalar Multiplication

Linear versus Affine Functions

Norms and Unit Vectors

Orthogonality

The Canonical Basis

Linear Independence and Dimension

## Matrices

Matrices and Their Transposes

Matrix Multiplication: Definition

## Vectors and Inner Products

Let  $\mathbf{x} = (x_i)_{i=1}^m \in \mathbb{R}^m$  denote a **column**  $m$ -vector of the form

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}.$$

Its **transpose** is the **row**  $m$ -vector

$$\mathbf{x}^\top = (x_1, x_2, \dots, x_m).$$

Given a column  $m$ -vector  $\mathbf{x}$  and row  $n$ -vector  $\mathbf{y}^\top = (y_j)_{j=1}^n \in \mathbb{R}^n$  where  $m = n$ , the **dot** or **scalar** or **inner product** is defined as

$$\mathbf{y}^\top \mathbf{x} := \mathbf{y} \cdot \mathbf{x} := \sum_{i=1}^n y_i x_i.$$

But when  $m \neq n$ , the scalar product is not defined.

# Exercise on Quadratic Forms

## Exercise

Consider the **quadratic form**  $f(\mathbf{w}) := \mathbf{w}^\top \mathbf{w}$   
as a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  of the column  $n$ -vector  $\mathbf{w}$ .

Explain why  $f(\mathbf{w}) \geq 0$  for all  $\mathbf{w} \in \mathbb{R}^n$ ,  
with equality if and only if  $\mathbf{w} = \mathbf{0}$ ,  
where  $\mathbf{0}$  denotes the zero vector of  $\mathbb{R}^n$ .

## Net Quantity Vectors

Suppose there are  $n$  commodities numbered from  $i = 1$  to  $n$ .

Each component  $q_i$  of the **net quantity vector**  $\mathbf{q} = (q_i)_{i=1}^n \in \mathbb{R}^n$  represents the quantity of the  $i$ th commodity.

Often each such quantity is non-negative.

But general equilibrium theory, following Debreu's *Theory of Value*, often uses only the sign of  $q_i$  to distinguish between

- ▶ a consumer's demands and supplies of the  $i$ th commodity;
- ▶ or a producer's outputs and inputs of the  $i$ th commodity.

This sign is taken to be

**positive** for demands or outputs;

**negative** for supplies or inputs.

In fact,  $q_i$  is taken to be

- ▶ the consumer's **net demand** for the  $i$ th commodity;
- ▶ the producer's **net supply** or **net output** of the  $i$ th commodity.

Then  $\mathbf{q}$  is the **net quantity vector**.

# Price Vectors

Each component  $p_i$  of the (row) **price vector**  $\mathbf{p}^\top \in \mathbb{R}^n$  indicates the price per unit of commodity  $i$ .

Then the scalar product

$$\mathbf{p}^\top \mathbf{q} = \mathbf{p} \cdot \mathbf{q} = \sum_{i=1}^n p_i q_i$$

is the total value of the net quantity vector  $\mathbf{q}$  evaluated at the price vector  $\mathbf{p}$ .

In particular,  $\mathbf{p}^\top \mathbf{q}$  indicates

- ▶ the net profit (or minus the net loss) for a producer;
- ▶ the net dissaving for a consumer.

# Outline

## Solving Two Equations in Two Unknowns

First Example

## Vectors

Vectors and Inner Products

**Addition, Subtraction, and Scalar Multiplication**

Linear versus Affine Functions

Norms and Unit Vectors

Orthogonality

The Canonical Basis

Linear Independence and Dimension

## Matrices

Matrices and Their Transposes

Matrix Multiplication: Definition



## Definitions

Consider any two  $n$ -vectors  $\mathbf{x} = (x_i)_{i=1}^n$  and  $\mathbf{y} = (y_i)_{i=1}^n$  in  $\mathbb{R}^n$ .

Their **sum**  $\mathbf{s} := \mathbf{x} + \mathbf{y}$  and **difference**  $\mathbf{d} := \mathbf{x} - \mathbf{y}$  are constructed by adding or subtracting the vectors component by component — i.e.,  $\mathbf{s} = (s_i)_{i=1}^n$  and  $\mathbf{d} = (d_i)_{i=1}^n$  where

$$s_i = x_i + y_i \quad \text{and} \quad d_i = x_i - y_i$$

for  $i = 1, 2, \dots, n$ .

The **scalar product**  $\lambda\mathbf{x}$  of any **scalar**  $\lambda \in \mathbb{R}$  and vector  $\mathbf{x} = (x_i)_{i=1}^n \in \mathbb{R}^n$  is constructed by multiplying each component of the vector  $\mathbf{x}$  by the scalar  $\lambda$  — i.e.,

$$\lambda\mathbf{x} = (\lambda x_i)_{i=1}^n$$

# Algebraic Fields

## Definition

An **algebraic field**  $(\mathbb{F}, +, \cdot)$  of scalars is a set  $\mathbb{F}$  that, together with the two **binary operations**  $+$  of **addition** and  $\cdot$  of **multiplication**, satisfies the following axioms for all  $a, b, c \in \mathbb{F}$ :

1.  $\mathbb{F}$  is **closed** under  $+$  and  $\cdot$ :  
— i.e., both  $a + b$  and  $a \cdot b$  are in  $\mathbb{F}$ .
2.  $+$  and  $\cdot$  are **associative**:  
both  $a + (b + c) = (a + b) + c$  and  $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ .
3.  $+$  and  $\cdot$  both **commute**:  
both  $a + b = b + a$  and  $a \cdot b = b \cdot a$ .
4. There are **identity** elements  $0, 1 \in \mathbb{F}$  for  $+$  and  $\cdot$  respectively, with  $0 \neq 1$ , such that: (i)  $a + 0 = a$ ; (ii)  $1 \cdot a = a$ .
5. There are **inverse** operations  $-$  for  $+$  and  $^{-1}$  for  $\cdot$  such that: (i)  $a + (-a) = 0$ ; (ii) provided  $a \neq 0$ , also  $a \cdot a^{-1} = 1$ .
6. The **distributive law**:  $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ .

# Three Examples of Real Algebraic Fields

## Exercise

Verify that the following well known sets are algebraic fields:

- ▶ the set  $\mathbb{R}$  of all real numbers,  
with the usual operations of addition and multiplication;
- ▶ the set  $\mathbb{Q}$  of all *rational numbers*  
— i.e., those that can be expressed as the ratio  $r = p/q$   
of integers  $p, q \in \mathbb{Z}$  with  $q \neq 0$ .  
(Check that  $\mathbb{Q}$  is closed  
under the usual operations of addition and multiplication,  
and that each non-zero rational  
has a rational multiplicative inverse.)
- ▶ the set  $\mathbb{Q} + \sqrt{2}\mathbb{Q} := \{r_1 + \sqrt{2}r_2 \mid r_1, r_2 \in \mathbb{Q}\} \subset \mathbb{R}$   
of all real numbers that can be expressed as the sum of:  
(i) a rational number;  
(ii) a rational multiple of the irrational number  $\sqrt{2}$ .

# Two Examples of Complex Algebraic Fields

## Exercise

Verify that the following well known sets are algebraic fields:

- ▶  $\mathbb{C}$ , the set of all **complex numbers**

— i.e., those that can be expressed as  $c = a + ib$ , where  $a, b \in \mathbb{R}$  and  $i$  is defined to satisfy  $i^2 = -1$ .

Note that  $\mathbb{C}$  is effectively the set  $\mathbb{R} \times \mathbb{R}$  of ordered pairs  $(a, b)$  satisfying  $a, b \in \mathbb{R}$ , together with the operations of:

(i) addition defined by  $(a, b) + (c, d) = (a + c, b + d)$

because  $(a + bi) + (c + di) = (a + c) + (b + d)i$ ;

(ii) multiplication defined by

$(a, b) \cdot (c, d) = (ac - bd, ad + bc)$

because  $(a + bi)(c + di) = (ac - bd) + (ad + bc)i$ .

- ▶ the set of all **rational complex numbers**

— i.e., those that can be expressed as  $c = a + ib$ , where  $a, b \in \mathbb{Q}$  and  $i$  is defined to satisfy  $i^2 = -1$ .

# General Vector Spaces

## Definition

A **vector** (or **linear**) space  $V$  over an algebraic field  $\mathbb{F}$  is a combination  $\langle V, \mathbb{F}, +, \cdot \rangle$  of:

- ▶ a set  $V$  of **vectors**;
- ▶ the field  $\mathbb{F}$  of **scalars**;
- ▶ the binary operation  $V \times V \ni (\mathbf{u}, \mathbf{v}) \mapsto \mathbf{u} + \mathbf{v} \in V$  of **vector addition**
- ▶ the binary operation  $\mathbb{F} \times V \ni (\alpha, \mathbf{u}) \mapsto \alpha\mathbf{u} \in V$  of **multiplication by a scalar**.

These are required to satisfy all of the following eight vector space axioms.

## Eight Vector Space Axioms

1. Addition is **associative**:  $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
2. Addition is **commutative**:  $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
3. Additive identity: There exists a **zero vector**  $\mathbf{0} \in V$  such that  $\mathbf{v} + \mathbf{0} = \mathbf{v}$  for all  $\mathbf{v} \in V$ .
4. Additive inverse: For every  $\mathbf{v} \in V$ , there exists an **additive inverse**  $-\mathbf{v} \in V$  of  $\mathbf{v}$  such that  $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$
5. Multiplication by a scalar is **distributive w.r.t. vector addition**:  $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$
6. Multiplication by a scalar is **distributive w.r.t. field addition**:  $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$
7. Multiplication by a scalar and field multiplication are **compatible**:  $\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v}$
8. The unit element  $1 \in \mathbb{F}$  is an **identity element** for scalar multiplication:  $1\mathbf{v} = \mathbf{v}$

# Multiplication by the Zero Scalar

## Exercise

*Prove that  $0\mathbf{v} = \mathbf{0}$  for all  $\mathbf{v} \in V$ .*

*Hint: Which three axioms justify the following chain of equalities*

$$0\mathbf{v} = [1 + (-1)]\mathbf{v} = 1\mathbf{v} + (-1)\mathbf{v} = \mathbf{v} - \mathbf{v} = \mathbf{0} ?$$

# A General Class of Finite Dimensional Vector Spaces

## Exercise

Given an arbitrary algebraic field  $\mathbb{F}$ , let  $\mathbb{F}^n$  denote the space of all lists  $\langle a_i \rangle_{i=1}^n$  of  $n$  elements  $a_i \in \mathbb{F}$  — i.e., the  $n$ -fold Cartesian product of  $\mathbb{F}$  with itself.

1. Show how to construct the respective binary operations

$$\mathbb{F}^n \times \mathbb{F}^n \ni (\mathbf{x}, \mathbf{y}) \mapsto \mathbf{x} + \mathbf{y} \in \mathbb{F}^n$$

$$\mathbb{F} \times \mathbb{F}^n \ni (\lambda, \mathbf{x}) \mapsto \lambda \mathbf{x} \in \mathbb{F}^n$$

*of addition and scalar multiplication*

*so that  $(\mathbb{F}^n, \mathbb{F}, +, \times)$  is a vector space.*

2. Show too that subtraction and division by a (non-zero) scalar can be defined by  $\mathbf{v} - \mathbf{w} = \mathbf{v} + (-1)\mathbf{w}$  and  $\mathbf{v}/\alpha = (1/\alpha)\mathbf{v}$ .



## Two Particular Finite Dimensional Vector Spaces

From now on we mostly consider **real vector spaces** over the real field  $\mathbb{R}$ , and especially the space  $(\mathbb{R}^n, \mathbb{R}, +, \times)$  of  **$n$ -vectors** over  $\mathbb{R}$ .

We will consider, however, the space  $(\mathbb{C}^n, \mathbb{C}, +, \times)$  of  **$n$ -vectors** over  $\mathbb{C}$  — the complex plane — when considering:

- ▶ eigenvalues and diagonalization of square matrices;
- ▶ systems of linear difference and differential equations;
- ▶ the characteristic function of a random variable.

# Outline

## Solving Two Equations in Two Unknowns

First Example

## Vectors

Vectors and Inner Products

Addition, Subtraction, and Scalar Multiplication

**Linear versus Affine Functions**

Norms and Unit Vectors

Orthogonality

The Canonical Basis

Linear Independence and Dimension

## Matrices

Matrices and Their Transposes

Matrix Multiplication: Definition

# Linear Combinations

## Definition

A **linear combination** of vectors is the weighted sum  $\sum_{h=1}^k \lambda_h \mathbf{x}^h$ , where  $\mathbf{x}^h \in V$  and  $\lambda_h \in \mathbb{F}$  for  $h = 1, 2, \dots, k$ .

## Exercise

*By induction on  $k$ , show that the vector space axioms imply that any linear combination of vectors in  $V$  must also belong to  $V$ .*

# Linear Functions

## Definition

A function  $V \ni \mathbf{u} \mapsto f(\mathbf{u}) \in \mathbb{F}$  is **linear** provided that

$$f(\lambda\mathbf{u} + \mu\mathbf{v}) = \lambda f(\mathbf{u}) + \mu f(\mathbf{v})$$

for every linear combination  $\lambda\mathbf{u} + \mu\mathbf{v}$  of two vectors  $\mathbf{u}, \mathbf{v} \in V$ , with  $\lambda, \mu \in \mathbb{F}$ .

## Exercise

*Prove that the function  $V \ni \mathbf{u} \mapsto f(\mathbf{u}) \in \mathbb{F}$  is linear if and only if both:*

- 1. for every vector  $\mathbf{v} \in V$  and scalar  $\lambda \in \mathbb{F}$  one has  $f(\lambda\mathbf{v}) = \lambda f(\mathbf{v})$ ;*
- 2. for every pair of vectors  $\mathbf{u}, \mathbf{v} \in V$  one has  $f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v})$ .*

# Key Properties of Linear Functions

## Exercise

Use induction on  $k$  to show that if the function  $f : V \rightarrow \mathbb{F}$  is linear, then

$$f\left(\sum_{h=1}^k \lambda_h \mathbf{x}^h\right) = \sum_{h=1}^k \lambda_h f(\mathbf{x}^h)$$

for all linear combinations  $\sum_{h=1}^k \lambda_h \mathbf{x}^h$  in  $V$  — i.e.,  $f$  **preserves linear combinations**.

## Exercise

In case  $V = \mathbb{R}^n$  and  $\mathbb{F} = \mathbb{R}$ , show that any linear function is **homogeneous of degree 1**, meaning that  $f(\lambda \mathbf{v}) = \lambda f(\mathbf{v})$  for all  $\lambda \in \mathbb{R}$  and all  $\mathbf{v} \in \mathbb{R}^n$ .

In particular, putting  $\lambda = 0$  gives  $f(\mathbf{0}) = 0$ .

What is the corresponding property in case  $V = \mathbb{Q}^n$  and  $\mathbb{F} = \mathbb{Q}$ ?

# Affine Functions

## Definition

A function  $g : V \rightarrow \mathbb{F}$  is said to be **affine** if there is a scalar **additive constant**  $\alpha \in \mathbb{F}$  and a linear function  $f : V \rightarrow \mathbb{F}$  such that  $g(\mathbf{v}) \equiv \alpha + f(\mathbf{v})$ .

## Exercise

*Under what conditions is an affine function  $g : \mathbb{R} \rightarrow \mathbb{R}$  linear when its domain  $\mathbb{R}$  is regarded as a vector space?*

# An Economic Aggregation Theorem

Suppose that a finite population of households  $h \in H$  with respective non-negative incomes  $y_h \in \mathbb{Q}_+$  ( $h \in H$ ) have non-negative demands  $x_h \in \mathbb{R}$  ( $h \in H$ ) which depend on household income via a function  $y_h \mapsto f_h(y_h)$ .

Given total income  $Y := \sum_h y_h$ , under what conditions can their total demand  $X := \sum_h x_h = \sum_h f_h(y_h)$  be expressed as a function  $X = F(Y)$  of  $Y$  alone?

The answer is an implication of **Cauchy's functional equation**.

In this context the theorem asserts that this **aggregation condition** implies that the functions  $f_h$  ( $h \in H$ ) and  $F$  must be **co-affine**.

This means there exists a **common** multiplicative constant  $\rho \in \mathbb{R}$ , along with additive constants  $\alpha_h$  ( $h \in H$ ) and  $A$ , such that

$$f_h(y_h) \equiv \alpha_h + \rho y_h \quad (h \in H) \quad \text{and} \quad F(Y) \equiv A + \rho Y$$

# Cauchy's Functional Equation: Proof of Sufficiency

## Theorem

Except in the trivial case when  $H$  has only one member, Cauchy's functional equation  $F(\sum_{h \in H} y_h) \equiv \sum_{h \in H} f_h(y_h)$  is satisfied for functions  $F, f_h : \mathbb{Q} \rightarrow \mathbb{R}$  if and only if:

1. there exists a single function  $\phi : \mathbb{Q} \rightarrow \mathbb{R}$  such that

$$F(q) = F(0) + \phi(q) \text{ and } f_h(q) = f_h(0) + \phi(q) \text{ for all } h \in H$$

2. the function  $\phi : \mathbb{Q} \rightarrow \mathbb{R}$  is linear, implying that the functions  $F$  and  $f_h$  are co-affine.

## Proof.

Suppose  $f_h(y_h) \equiv \alpha_h + \rho y_h$  for all  $h \in H$ , and  $F(Y) \equiv A + \rho Y$ . Then Cauchy's functional equation  $F(\sum_{h \in H} y_h) \equiv \sum_{h \in H} f_h(y_h)$  is obviously satisfied provided that  $A = \sum_{h \in H} \alpha_h$ . □



# Cauchy's Equation: Beginning the Proof of Necessity

## Lemma

The mapping  $\mathbb{Q} \ni q \mapsto \phi(q) := F(q) - F(0) \in \mathbb{R}$  must satisfy;

1.  $\phi(q) \equiv f_i(q) - f_i(0)$  for all  $i \in H$  and  $q \in \mathbb{Q}$ ;
2.  $\phi(q + q') \equiv \phi(q) + \phi(q')$  for all  $q, q' \in \mathbb{Q}$ .

## Proof.

To prove part 1, consider any  $i \in H$  and all  $q \in \mathbb{Q}$ .

Note that Cauchy's equation  $F(\sum_h y_h) \equiv \sum_h f_h(y_h)$

implies that  $F(q) = f_i(q) + \sum_{h \neq i} f_h(0)$

and also  $F(0) = f_i(0) + \sum_{h \neq i} f_h(0)$ .

Now define the function  $\phi(q) := F(q) - F(0)$  on the domain  $\mathbb{Q}$ .

Then subtract the second equation from the first to obtain

$$\phi(q) = F(q) - F(0) = f_i(q) - f_i(0)$$



# Cauchy's Equation: Continuing the Proof of Necessity

Proof.

To prove part 2, consider any  $i, j \in H$  with  $i \neq j$ , and any  $q, q' \in \mathbb{Q}$ .

Note that Cauchy's equation  $F(\sum_h y_h) \equiv \sum_h f_h(y_h)$  implies that

$$\begin{aligned}F(q + q') &= f_i(q) + f_j(q') + \sum_{h \in H \setminus \{i, j\}} f_h(0) \\F(0) &= f_i(0) + f_j(0) + \sum_{h \in H \setminus \{i, j\}} f_h(0)\end{aligned}$$

Now subtract the second equation from the first, then use the equation  $\phi(q) = F(q) - F(0) = f_i(q) - f_i(0)$  derived in the previous slide, to obtain successively

$$\begin{aligned}\phi(q + q') &= F(q + q') - F(0) \\&= f_i(q) - f_i(0) + f_j(q') - f_j(0) \\&= \phi(q) + \phi(q')\end{aligned}$$



## Cauchy's Equation: Resuming the Proof of Necessity

Because  $\phi(q + q') \equiv \phi(q) + \phi(q')$ ,

for any  $k \in \mathbb{N}$  one has  $\phi(kq) = \phi((k-1)q) + \phi(q)$ .

As an induction hypothesis, which is trivially true for  $k = 2$ , suppose that  $\phi((k-1)q) = (k-1)\phi(q)$ .

Confirming the induction step, the hypothesis implies that

$$\phi(kq) = \phi((k-1)q) + \phi(q) = (k-1)\phi(q) + \phi(q) = k\phi(q)$$

So  $\phi(kq) = k\phi(q)$  for every  $k \in \mathbb{N}$  and every  $q \in \mathbb{Q}$ .

Putting  $q' = kq$  implies that  $\phi(q') = k\phi(q'/k)$ .

Interchanging  $q$  and  $q'$ , it follows that  $\phi(q/k) = (1/k)\phi(q)$ .

## Cauchy's Equation: Completing the Proof of Necessity

So far we have proved that, for every  $k \in \mathbb{N}$  and every  $q \in \mathbb{Q}$ , one has both  $\phi(kq) = k\phi(q)$  and  $\phi(q/k) = (1/k)\phi(q)$ .

Hence, for every rational  $r = m/n \in \mathbb{Q}$  one has  $\phi(mq/n) = m\phi(q/n) = (m/n)\phi(q)$  and so  $\phi(rq) = r\phi(q)$ .

In particular,  $\phi(r) = r\phi(1)$ , so  $\phi$  is linear on its domain  $\mathbb{Q}$  (though not on the whole of  $\mathbb{R}$  without additional assumptions such as continuity or monotonicity).

The rest of the proof is routine checking of definitions.

# Outline

## Solving Two Equations in Two Unknowns

First Example

## Vectors

Vectors and Inner Products

Addition, Subtraction, and Scalar Multiplication

Linear versus Affine Functions

**Norms and Unit Vectors**

Orthogonality

The Canonical Basis

Linear Independence and Dimension

## Matrices

Matrices and Their Transposes

Matrix Multiplication: Definition

## Euclidean Norm as Length

Pythagoras's theorem implies that the **length** of the typical vector  $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$  is  $\sqrt{x_1^2 + x_2^2}$  or, perhaps less clumsily,  $(x_1^2 + x_2^2)^{1/2}$ .

In  $\mathbb{R}^3$ , the same result implies that the **length** of the typical vector  $\mathbf{x} = (x_1, x_2, x_3)$  is

$$\left[ \left( (x_1^2 + x_2^2)^{1/2} \right)^2 + x_3^2 \right]^{1/2} = (x_1^2 + x_2^2 + x_3^2)^{1/2}.$$

An obvious extension to  $\mathbb{R}^n$  is the following:

### Definition

The **length** of the typical  $n$ -vector  $\mathbf{x} = (x_i)_{i=1}^n \in \mathbb{R}^n$  is its **(Euclidean) norm**

$$\|\mathbf{x}\| := \left( \sum_{i=1}^n x_i^2 \right)^{1/2} = \sqrt{\mathbf{x}^\top \mathbf{x}} = \sqrt{\mathbf{x} \cdot \mathbf{x}}$$

# Unit $n$ -Vectors, the Unit Sphere, and Unit Ball

## Definition

A **unit** vector  $\mathbf{u} \in \mathbb{R}^n$  is a vector with unit norm — i.e., its components satisfy  $\sum_{i=1}^n u_i^2 = \|\mathbf{u}\|^2 = 1$ .

The set of all such unit vectors forms a surface called the **unit sphere** of dimension  $n - 1$  (one less than  $n$  because of the defining equation).

It is defined as the hollow set (like a football or tennis ball)

$$S^{n-1} := \left\{ \mathbf{x} \in \mathbb{R}^n \mid \sum_{i=1}^n x_i^2 = 1 \right\}$$

The **unit ball**  $B \subset \mathbb{R}^n$  is the solid set (like a cricket ball or golf ball)

$$B := \left\{ \mathbf{x} \in \mathbb{R}^n \mid \sum_{i=1}^n x_i^2 \leq 1 \right\}$$

of all points bounded by the surface of the unit sphere  $S^{n-1} \subset \mathbb{R}^n$ .

# Cauchy–Schwartz Inequality

## Theorem

For all pairs  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ , one has  $|\mathbf{a} \cdot \mathbf{b}| \leq \|\mathbf{a}\| \|\mathbf{b}\|$ .

## Proof.

Define the function  $\mathbb{R} \ni \xi \mapsto f(\xi) := \sum_{i=1}^n (a_i \xi + b_i)^2 \in \mathbb{R}$ .

Clearly  $f$  is the quadratic function  $f(\xi) \equiv A\xi^2 + B\xi + C$

where  $A := \sum_{i=1}^n a_i^2 = \|\mathbf{a}\|^2$ ,  $B := 2 \sum_{i=1}^n a_i b_i = 2\mathbf{a} \cdot \mathbf{b}$ ,

and  $C := \sum_{i=1}^n b_i^2 = \|\mathbf{b}\|^2$ .

There is a trivial case when  $A = 0$  because  $\mathbf{a} = \mathbf{0}$ .

Otherwise  $A > 0$ , so we can complete the square to get

$$f(\xi) \equiv A\xi^2 + B\xi + C = A[\xi + (B/2A)]^2 + C - B^2/4A$$

But the definition of  $f$  implies that  $f(\xi) \geq 0$  for all  $\xi \in \mathbb{R}$ , including  $\xi = -B/2A$ , so  $0 \leq f(-B/2A) = C - B^2/4A$ .

Hence  $\frac{1}{4}B^2 \leq AC$ ,

implying that  $|\mathbf{a} \cdot \mathbf{b}| = \left| \frac{1}{2}B \right| \leq \sqrt{AC} = \|\mathbf{a}\| \|\mathbf{b}\|$ . □



# Outline

## Solving Two Equations in Two Unknowns

First Example

## Vectors

Vectors and Inner Products

Addition, Subtraction, and Scalar Multiplication

Linear versus Affine Functions

Norms and Unit Vectors

**Orthogonality**

The Canonical Basis

Linear Independence and Dimension

## Matrices

Matrices and Their Transposes

Matrix Multiplication: Definition

# The Angle Between Two Vectors

Consider the triangle in  $\mathbb{R}^n$  whose vertices are the three disjoint vectors  $\mathbf{x}$ ,  $\mathbf{y}$ ,  $\mathbf{0}$ .

Its three sides or edges have respective lengths  $\|\mathbf{x}\|$ ,  $\|\mathbf{y}\|$ ,  $\|\mathbf{x} - \mathbf{y}\|$ , where the last follows from the parallelogram law.

Note that  $\|\mathbf{x} - \mathbf{y}\|^2 \begin{matrix} \leq \\ > \end{matrix} \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$  according as the angle at  $\mathbf{0}$  is: (i) acute; (ii) a right angle; (iii) obtuse. But

$$\begin{aligned}\|\mathbf{x} - \mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2 &= \sum_{i=1}^n (x_i - y_i)^2 - \sum_{i=1}^n (x_i^2 + y_i^2) \\ &= \sum_{i=1}^n (-2x_i y_i) = -2\mathbf{x} \cdot \mathbf{y}\end{aligned}$$

So the three cases (i)–(iii) occur according as  $\mathbf{x} \cdot \mathbf{y} \begin{matrix} \geq \\ < \end{matrix} 0$ .

Using the Cauchy–Schwartz inequality, one can define the **angle** between  $\mathbf{x}$  and  $\mathbf{y}$  as the unique solution  $\theta = \arccos(\mathbf{x} \cdot \mathbf{y} / \|\mathbf{x}\| \|\mathbf{y}\|)$  in the interval  $[0, \pi)$  of the equation  $\cos \theta = \mathbf{x} \cdot \mathbf{y} / \|\mathbf{x}\| \|\mathbf{y}\| \in [-1, 1]$ .

## Orthogonal and Orthonormal Sets of Vectors

Case (ii) suggests defining two vectors  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

as **orthogonal** just in case  $\mathbf{x} \cdot \mathbf{y} = 0$ ,

which is true if and only if  $\theta = \arccos 0 = \frac{1}{2}\pi = 90^\circ$ .

A set of  $k$  vectors  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\} \subset \mathbb{R}^n$  is said to be:

- ▶ **pairwise orthogonal** just in case  $\mathbf{x}_i \cdot \mathbf{x}_j = 0$  whenever  $j \neq i$ ;
- ▶ **orthonormal** just in case, in addition, each  $\|\mathbf{x}_i\| = 1$   
— i.e., all  $k$  elements of the set are vectors of unit length.

Define the **Kronecker delta** function

$$\{1, 2, \dots, n\} \times \{1, 2, \dots, n\} \ni (i, j) \mapsto \delta_{ij} \in \{0, 1\}$$

on the set of pairs  $i, j \in \{1, 2, \dots, n\}$  by

$$\delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Then the set of  $k$  vectors  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\} \subset \mathbb{R}^n$  is orthonormal if and only if  $\mathbf{x}_i \cdot \mathbf{x}_j = \delta_{ij}$  for all pairs  $i, j \in \{1, 2, \dots, k\}$ .

# Outline

## Solving Two Equations in Two Unknowns

First Example

## Vectors

Vectors and Inner Products

Addition, Subtraction, and Scalar Multiplication

Linear versus Affine Functions

Norms and Unit Vectors

Orthogonality

**The Canonical Basis**

Linear Independence and Dimension

## Matrices

Matrices and Their Transposes

Matrix Multiplication: Definition

# The Canonical Basis of $\mathbb{R}^n$

## Example

A prominent orthonormal set is the **canonical basis** of  $\mathbb{R}^n$ , defined as the set of  $n$  different  $n$ -vectors  $\mathbf{e}^i$  ( $i = 1, 2, \dots, n$ ) whose respective components  $(e_j^i)_{j=1}^n$  satisfy  $e_j^i = \delta_{ij}$  for all  $j \in \{1, 2, \dots, n\}$ .

## Exercise

Show that each  $n$ -vector  $\mathbf{x} = (x_i)_{i=1}^n$  is a linear combination

$$\mathbf{x} = (x_i)_{i=1}^n = \sum_{i=1}^n x_i \mathbf{e}^i$$

of the canonical basis vectors  $\{\mathbf{e}^1, \mathbf{e}^2, \dots, \mathbf{e}^n\}$ , with the multiplier attached to each basis vector  $\mathbf{e}^i$  equal to the respective component  $x_i$  ( $i = 1, 2, \dots, n$ ).

# The Canonical Basis in Commodity Space

## Example

Consider the case when each vector  $\mathbf{x} \in \mathbb{R}^n$  is a **quantity vector**, whose components are  $(x_i)_{i=1}^n$ , where  $x_i$  indicates the net quantity of commodity  $i$ .

Then the  $i$ th unit vector  $\mathbf{e}^i$  of the canonical basis of  $\mathbb{R}^n$  represents a **commodity bundle** that consists of one unit of commodity  $i$ , but nothing of every other commodity.

In case the row vector  $\mathbf{p}^\top \in \mathbb{R}^n$  is a price vector for the same list of  $n$  commodities, the value  $\mathbf{p}^\top \mathbf{e}^i$  of the  $i$ th unit vector  $\mathbf{e}^i$  must equal  $p_i$ , the price (of one unit) of the  $i$ th commodity.

# Linear Functions

## Theorem

*The function  $\mathbb{R}^n \ni \mathbf{x} \mapsto f(\mathbf{x}) \in \mathbb{R}$  is linear if and only if there exists  $\mathbf{y} \in \mathbb{R}^n$  such that  $f(\mathbf{x}) = \mathbf{y}^\top \mathbf{x}$ .*

## Proof.

Sufficiency is easy to check.

Conversely, note that  $\mathbf{x}$  equals the linear combination  $\sum_{i=1}^n x_i \mathbf{e}^i$  of the  $n$  canonical basis vectors.

Hence, linearity of  $f$  implies that

$$f(\mathbf{x}) = f\left(\sum_{i=1}^n x_i \mathbf{e}^i\right) = \sum_{i=1}^n x_i f(\mathbf{e}^i) = \sum_{i=1}^n f(\mathbf{e}^i) x_i = \mathbf{y}^\top \mathbf{x}$$

where  $\mathbf{y}$  is the column vector whose components are  $y_i = f(\mathbf{e}^i)$  for  $i = 1, 2, \dots, n$ . □

# Linear Transformations: Definition

## Definition

The vector-valued function

$$\mathbb{R}^n \ni \mathbf{x} \mapsto \mathbf{F}(\mathbf{x}) = (F_i(\mathbf{x}))_{i=1}^m \in \mathbb{R}^m$$

is a **linear transformation** just in case

each component function  $\mathbb{R}^n \ni \mathbf{x} \mapsto F_i(\mathbf{x}) \in \mathbb{R}$  is linear

— or equivalently, iff  $\mathbf{F}(\lambda\mathbf{x} + \mu\mathbf{y}) = \lambda\mathbf{F}(\mathbf{x}) + \mu\mathbf{F}(\mathbf{y})$

for every linear combination  $\lambda\mathbf{x} + \mu\mathbf{y}$  of every pair  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ .



# Characterizing Linear Transformations

## Theorem

*The mapping  $\mathbb{R}^n \ni \mathbf{x} \mapsto \mathbf{F}(\mathbf{x}) \in \mathbb{R}^m$  is a linear transformation if and only if there exist vectors  $\mathbf{y}_i \in \mathbb{R}^m$  for  $i = 1, 2, \dots, n$  such that each component function satisfies  $F_i(\mathbf{x}) = \mathbf{y}_i^\top \mathbf{x}$ .*

## Proof.

Sufficiency is obvious.

Conversely, because  $\mathbf{x}$  equals the linear combination  $\sum_{i=1}^n x_i \mathbf{e}_i$  of the  $n$  canonical basis vectors  $\{\mathbf{e}_i\}_{i=1}^n$  and because each component function  $\mathbb{R}^n \ni \mathbf{x} \mapsto F_i(\mathbf{x})$  is linear, one has

$$F_i(\mathbf{x}) = F_i\left(\sum_{j=1}^n x_j \mathbf{e}_j\right) = \sum_{j=1}^n x_j F_i(\mathbf{e}_j) = \mathbf{y}_i^\top \mathbf{x}$$

where  $\mathbf{y}_i^\top$  is the row vector whose components are  $(\mathbf{y}_i)_j = F_i(\mathbf{e}_j)$  for  $i = 1, 2, \dots, m$  and  $j = 1, 2, \dots, n$ . □

# Representing a Linear Transformation

## Definition

A **matrix representation**

of the linear transformation  $\mathbb{R}^n \ni \mathbf{x} \mapsto \mathbf{F}(\mathbf{x}) \in \mathbb{R}^m$

relative to the canonical bases of  $\mathbb{R}^n$  and  $\mathbb{R}^m$

is an  $m \times n$  array whose  $n$  columns

are the  $m$ -vector images  $\mathbf{F}(\mathbf{e}_j) = (F_i(\mathbf{e}_j))_{i=1}^m \in \mathbb{R}^m$

of the  $n$  canonical basis vectors  $\{\mathbf{e}_j\}_{j=1}^n$  of  $\mathbb{R}^n$ .

# Outline

## Solving Two Equations in Two Unknowns

First Example

## Vectors

Vectors and Inner Products

Addition, Subtraction, and Scalar Multiplication

Linear versus Affine Functions

Norms and Unit Vectors

Orthogonality

The Canonical Basis

**Linear Independence and Dimension**

## Matrices

Matrices and Their Transposes

Matrix Multiplication: Definition

# Linear Combinations and Dependence: Definitions

## Definition

A **linear combination** of the finite set  $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^k\}$  of vectors is the scalar weighted sum  $\sum_{h=1}^k \lambda_h \mathbf{x}^h$ , where  $\lambda_h \in \mathbb{R}$  for  $h = 1, 2, \dots, k$ .

## Definition

The finite set  $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^k\}$  of vectors is **linearly independent** just in case the only solution of the equation  $\sum_{h=1}^k \lambda_h \mathbf{x}^h = \mathbf{0}$  is the **trivial solution**  $\lambda_1 = \lambda_2 = \dots = \lambda_k = 0$ .

Alternatively, if the equation has a non-trivial solution, then the set of vectors is **linearly dependent**.

# Characterizing Linear Dependence

## Theorem

The finite set  $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^k\}$  of  $k$   $n$ -vectors is linearly dependent if and only if at least one of the vectors, say  $\mathbf{x}^1$  after reordering, can be expressed as a linear combination of the others — i.e., there exist scalars  $\alpha^h$  ( $h = 2, 3, \dots, k$ ) such that  $\mathbf{x}^1 = \sum_{h=2}^k \alpha_h \mathbf{x}^h$ .

## Proof.

If  $\mathbf{x}^1 = \sum_{h=2}^k \alpha_h \mathbf{x}^h$ , then  $(-1)\mathbf{x}^1 + \sum_{h=2}^k \alpha_h \mathbf{x}^h = \mathbf{0}$ , so  $\sum_{h=1}^k \lambda_h \mathbf{x}^h = \mathbf{0}$  has a non-trivial solution.

Conversely, suppose  $\sum_{h=1}^k \lambda_h \mathbf{x}^h = \mathbf{0}$  has a non-trivial solution.

After reordering, we can suppose that  $\lambda_1 \neq 0$ .

Then  $\mathbf{x}^1 = \sum_{h=2}^k \alpha_h \mathbf{x}^h$ ,

where  $\alpha_h = -\lambda_h/\lambda_1$  for  $h = 2, 3, \dots, k$ . □

# Orthogonality Implies Linear Independence

## Theorem

*If the finite set  $S = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^k\}$  of  $k$  non-zero  $n$ -vectors is pairwise orthogonal, then it is linearly independent.*

## Proof.

Let  $\mathbf{s}$  denote the linear combination  $\sum_{h=1}^k \alpha_h \mathbf{x}^h$ .

Then for each  $j = 1, \dots, k$  one has  $\mathbf{s} \cdot \mathbf{x}^j = \sum_{h=1}^k \alpha_h \mathbf{x}^h \cdot \mathbf{x}^j$ .

In case  $S$  is pairwise orthogonal, one has  $\mathbf{x}^h \cdot \mathbf{x}^j = 0$  for all  $h \neq j$ , and so  $\mathbf{s} \cdot \mathbf{x}^j = \alpha_j \mathbf{x}^j \cdot \mathbf{x}^j$ .

So when  $\mathbf{s} = \mathbf{0}$ , it follows that  $\alpha_j \mathbf{x}^j \cdot \mathbf{x}^j = 0$  for all  $j = 1, \dots, k$ .

Then, because we assumed that  $\mathbf{x}^j \neq \mathbf{0}$ , we have  $\alpha_j = 0$  for all  $j = 1, \dots, k$ .

This proves that  $S$  is linearly independent. □

# Dimension

## Definition

The **dimension** of a vector space  $V$  is the size of any maximal set of linearly independent vectors, if this number is finite.

Otherwise, if there is an infinite set of linearly independent vectors, the dimension is **infinite**.

## Exercise

*Show that the canonical basis of  $\mathbb{R}^n$  is linearly independent.*

## Example

The previous exercise shows that the dimension of  $\mathbb{R}^n$  is at least  $n$ .

Later results will imply that any set of  $k > n$  vectors in  $\mathbb{R}^n$  is linearly dependent.

This implies that the dimension of  $\mathbb{R}^n$  is exactly  $n$ .

# Outline

## Solving Two Equations in Two Unknowns

First Example

## Vectors

Vectors and Inner Products

Addition, Subtraction, and Scalar Multiplication

Linear versus Affine Functions

Norms and Unit Vectors

Orthogonality

The Canonical Basis

Linear Independence and Dimension

## Matrices

**Matrices and Their Transposes**

Matrix Multiplication: Definition



## Matrices as Rectangular Arrays

An  $m \times n$  **matrix**  $\mathbf{A} = (a_{ij})_{m \times n}$  is a (rectangular) array

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = ((a_{ij})_{i=1}^m)_{j=1}^n = ((a_{ij})_{j=1}^n)_{i=1}^m$$

Note that in  $a_{ij}$ , we write the **row** number  $i$  **before** the **column** number  $j$ .

An  $m \times 1$  matrix is a **column vector** with  $m$  rows and 1 column.

A  $1 \times n$  matrix is a **row vector** with 1 row and  $n$  columns.

The  $m \times n$  **matrix**  $\mathbf{A}$  consists of:

$n$  **columns** in the form of  $m$ -vectors

$$\mathbf{a}_j = (a_{ij})_{i=1}^m \in \mathbb{R}^m \text{ for } j = 1, 2, \dots, n;$$

$m$  **rows** in the form of  $n$ -vectors

$$\mathbf{a}_i^\top = (a_{ij})_{j=1}^n \in \mathbb{R}^n \text{ for } i = 1, 2, \dots, m.$$

# The Transpose of a Matrix

The **transpose** of the  $m \times n$  matrix  $\mathbf{A} = (a_{ij})_{m \times n}$  is defined as the  $n \times m$  matrix

$$\mathbf{A}^{\top} = (a_{ij}^{\top})_{n \times m} = (a_{ji})_{n \times m} = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{pmatrix}$$

Thus the transposed matrix  $\mathbf{A}^{\top}$  results from transforming each column  $m$ -vector  $\mathbf{a}_j = (a_{ij})_{i=1}^m$  ( $j = 1, 2, \dots, n$ ) of  $\mathbf{A}$  into the corresponding row  $m$ -vector  $\mathbf{a}_j^{\top} = (a_{ji}^{\top})_{i=1}^m$  of  $\mathbf{A}^{\top}$ .

Equivalently, for each  $i = 1, 2, \dots, m$ , the  $i$ th row  $n$ -vector  $\mathbf{a}_i^{\top} = (a_{ij})_{j=1}^n$  of  $\mathbf{A}$  is transformed into the  $i$ th column  $n$ -vector  $\mathbf{a}_i = (a_{ji})_{j=1}^n$  of  $\mathbf{A}^{\top}$ .

Either way, one has  $a_{ij}^{\top} = a_{ji}$  for all relevant pairs  $i, j$ .

# Rows Before Columns

**VERY Important Rule:** Rows **before** columns!

This order really matters.

Reversing it gives a transposed matrix.

## Exercise

Verify that the double transpose of any  $m \times n$  matrix  $\mathbf{A}$  satisfies  $(\mathbf{A}^\top)^\top = \mathbf{A}$

— i.e., transposing a matrix twice recovers the original matrix.

# Outline

## Solving Two Equations in Two Unknowns

First Example

## Vectors

Vectors and Inner Products

Addition, Subtraction, and Scalar Multiplication

Linear versus Affine Functions

Norms and Unit Vectors

Orthogonality

The Canonical Basis

Linear Independence and Dimension

## Matrices

Matrices and Their Transposes

**Matrix Multiplication: Definition**

# Multiplying a Matrix by a Scalar

A **scalar**, usually denoted by a Greek letter, is simply a member  $\alpha \in \mathbb{F}$  of the algebraic field  $\mathbb{F}$  over which the vector space is defined.

So when  $\mathbb{F} = \mathbb{R}$ , a scalar is a real number  $\alpha \in \mathbb{R}$ .

The **product** of any  $m \times n$  matrix  $\mathbf{A} = (a_{ij})_{m \times n}$  and any scalar  $\alpha \in \mathbb{R}$  is the new  $m \times n$  matrix denoted by  $\alpha\mathbf{A} = (\alpha a_{ij})_{m \times n}$ , each of whose elements  $\alpha a_{ij}$  results from multiplying the corresponding element  $a_{ij}$  of  $\mathbf{A}$  by  $\alpha$ .

# Matrix Multiplication

The **matrix product** of two matrices **A** and **B** is defined (whenever possible) as the matrix  $\mathbf{C} = \mathbf{AB} = (c_{ij})_{m \times n}$  whose element  $c_{ij}$  in row  $i$  and column  $j$  is the inner product  $c_{ij} = \mathbf{a}_i^\top \mathbf{b}_j$  of:

- ▶ the  $i$ th **row** vector  $\mathbf{a}_i^\top$  of the first matrix **A**;
- ▶ the  $j$ th **column** vector  $\mathbf{b}_j$  of the second matrix **B**.

$$\begin{pmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ij} & \dots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mj} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & \dots & b_{1j} & \dots & b_{1p} \\ \vdots & & \vdots & & \vdots \\ b_{j1} & \dots & b_{jj} & \dots & b_{jp} \\ \vdots & & \vdots & & \vdots \\ b_{m1} & \dots & b_{mj} & \dots & b_{mp} \end{pmatrix} = \begin{pmatrix} c_{11} & \dots & c_{1j} & \dots & c_{1p} \\ \vdots & & \vdots & & \vdots \\ c_{i1} & \dots & c_{ij} & \dots & c_{ip} \\ \vdots & & \vdots & & \vdots \\ c_{m1} & \dots & c_{mj} & \dots & c_{mp} \end{pmatrix}$$

$\mathbf{a}_i^\top \quad \cdot \quad \mathbf{b}_j \quad = \quad c_{ij}$

# Compatibility for Matrix Multiplication, I

Again: rows **before** columns!

Note that the resulting matrix product **C** must have:

- ▶ as many rows as the first matrix **A**;
- ▶ as many columns as the second matrix **B**.

Yet again: rows **before** columns!

## Compatibility for Matrix Multiplication, II

**Question:** when is this definition  
of the matrix product  $\mathbf{C} = \mathbf{AB}$  possible?

**Answer:** if and only if  $\mathbf{A}$  has as many columns as  $\mathbf{B}$  has rows.

This condition ensures that every inner product  $\mathbf{a}_i^\top \mathbf{b}_j$  is defined, which is true iff (if and only if) every row of  $\mathbf{A}$  has exactly the same number of elements as every column of  $\mathbf{B}$ .

In this case, the two matrices  $\mathbf{A}$  and  $\mathbf{B}$  are **compatible for multiplication**.

Specifically, if  $\mathbf{A}$  is  $m \times \ell$  for some  $m$ , then  $\mathbf{B}$  must be  $\ell \times n$  for some  $n$ .

Then the product  $\mathbf{C} = \mathbf{AB}$  is  $m \times n$ , with elements  $c_{ij} = \mathbf{a}_i^\top \mathbf{b}_j = \sum_{k=1}^{\ell} a_{ik} b_{kj}$  for  $i = 1, 2, \dots, m$  and  $j = 1, 2, \dots, n$ .



# Laws of Matrix Multiplication

## Exercise

Verify that the following *laws of matrix multiplication* hold whenever the matrices  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  are compatible for multiplication.

*associative law for matrices:*  $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$ ;

*distributive:*  $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$  and  $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$ ;

*transpose:*  $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$ .

*associative law for scalars:*  $\alpha(\mathbf{AB}) = (\alpha\mathbf{A})\mathbf{B} = \mathbf{A}(\alpha\mathbf{B})$  (all  $\alpha \in \mathbb{R}$ ).

## Exercise

Let  $\mathbf{X}$  be any  $m \times n$  matrix, and  $\mathbf{z}$  any column  $n$ -vector.

1. Show that the matrix product  $\mathbf{z}^\top \mathbf{X}^\top \mathbf{X} \mathbf{z}$  is well-defined, and that its value is a scalar.
2. By putting  $\mathbf{w} = \mathbf{X} \mathbf{z}$  in the previous exercise regarding the sign of the quadratic form  $\mathbf{w}^\top \mathbf{w}$ , what can you conclude about the value of the scalar  $\mathbf{z}^\top \mathbf{X}^\top \mathbf{X} \mathbf{z}$ ?

# Exercise for Econometricians I

## Exercise

An econometrician has access to data series (such as time series) involving the real values

- ▶  $y_t$  ( $t = 1, 2, \dots, T$ ) of one *endogenous* variable;
- ▶  $x_{ti}$  ( $t = 1, 2, \dots, T$  and  $i = 1, 2, \dots, k$ ) of  $k$  different *exogenous* variables  
— sometimes called *explanatory* variables or *regressors*.

The data is to be fitted into the *linear regression model*

$$y_t = \sum_{i=1}^k b_i x_{ti} + e_t$$

whose scalar constants  $b_i$  ( $i = 1, 2, \dots, k$ ) are unknown *regression coefficients*, and each scalar  $e_t$  is the *error term* or *residual*.

## Exercise for Econometricians II

1. Discuss how the regression model can be written in the form  $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$  for suitable column vectors  $\mathbf{y}$ ,  $\mathbf{b}$ ,  $\mathbf{e}$ .
2. What are the dimensions of these vectors, and of the exogenous data matrix  $\mathbf{X}$ ?
3. Why do you think econometricians use this matrix equation, rather than the alternative  $\mathbf{y} = \mathbf{b}\mathbf{X} + \mathbf{e}$ ?
4. How can the equation  $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$  accommodate the constant term  $\alpha$  in the alternative equation  $y_t = \alpha + \sum_{i=1}^k b_i x_{ti} + e_t$ ?

## Matrix Multiplication Does Not Commute I

The two matrices **A** and **B** **commute** just in case **AB = BA**.

Note that typical pairs of matrices **DO NOT** commute, meaning that **AB ≠ BA** — i.e., the order of multiplication matters.

Indeed, suppose that **A** is  $\ell \times m$  and **B** is  $m \times n$ , as is needed for **AB** to be defined.

Then the reverse product **BA** is **undefined** except in the special case when  $n = \ell$ .

Hence, for both **AB** and **BA** to be defined, where **B** is  $m \times n$ , the matrix **A** **must** be  $n \times m$ .

But then **AB** is  $n \times n$ , whereas **BA** is  $m \times m$ .

Evidently **AB ≠ BA** unless  $m = n$ .

Thus all four matrices **A**, **B**, **AB** and **BA** are  $m \times m = n \times n$ .

We must be in the special case when all four are **square** matrices of the **same** dimension.

## Matrix Multiplication Does Not Commute II

Even if both **A** and **B** are  $n \times n$  matrices, implying that both **AB** and **BA** are also  $n \times n$ , one can still have **AB**  $\neq$  **BA**.

### Example

Here is a  $2 \times 2$  example:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

### Exercise

*For matrix multiplication, explain why there are two different versions of the distributive law — namely*

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC} \text{ and } (\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$$

# More Warnings Regarding Matrix Multiplication

## Exercise

Let  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  denote three general matrices.

Give examples showing that:

1. The matrix  $\mathbf{AB}$  might be defined, even if  $\mathbf{BA}$  is not.
2. One can have  $\mathbf{AB} = \mathbf{0}$  even though  $\mathbf{A} \neq \mathbf{0}$  and  $\mathbf{B} \neq \mathbf{0}$ .
3. If  $\mathbf{AB} = \mathbf{AC}$  and  $\mathbf{A} \neq \mathbf{0}$ , it does not follow that  $\mathbf{B} = \mathbf{C}$ .