Advanced Econometric Methods EMET3011/8014

Lecture 8

John Stachurski

Semester 1, 2011

Announcements/Reminders

- Please get yourself fresh copy of the course notes PDF
- Discussion of midterm exam at the end of the lecture
- No computer lab (or lecture) next week

Assignment 1

Solutions up on the web—please read

General comment: Nice work

Code:

- Be clear but concise—cut and paste is a bad sign
- Avoid "magic numbers"—replace with named constants

Discussion of results:

- Link carefully with theory—check the conditions

Today's Lecture

- Linear Functions
- Linear Subspaces
- Linear Independence
- Systems of Equations
- Random Matrices
- The OPT

Matrices as Maps

One way to view matrices: mappings from one space to another

Special property of these functions: They are linear

Def. Function $f\colon \mathbb{R}^K \to \mathbb{R}^N$ called **linear** if

$$f(\alpha \mathbf{x} + \beta \mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y}) \text{ for any } \mathbf{x}, \mathbf{y} \in \mathbb{R}^K \text{ and } \alpha, \beta \in \mathbb{R}$$

Example Function $f(x) = x^2$ is nonlinear, because if we take $\alpha = \beta = x = y = 1$, then

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

Fix $N \times K$ matrix $\mathbf{A}$

The function f defined by $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is linear

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

The rules of matrix addition and scalar multiplication yield

$$\begin{aligned} f(\alpha\mathbf{x} + \beta\mathbf{y}) &:= \mathbf{A}(\alpha\mathbf{x} + \beta\mathbf{y}) \\ &= \mathbf{A}\alpha\mathbf{x} + \mathbf{A}\beta\mathbf{y} \\ &= \alpha\mathbf{A}\mathbf{x} + \beta\mathbf{A}\mathbf{y} \\ &=: \alpha f(\mathbf{x}) + \beta f(\mathbf{y}) \end{aligned}$$

In other words, f is linear

How about some more examples of linear functions?

There aren't any

The next theorem summarizes

Theorem.

1. If $\mathbf{A}$ is an $N \times K$ matrix and $f \colon R^K \to \mathbb{R}^N$ is defined by $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$, then f is linear.
2. Conversely, if $f \colon \mathbb{R}^K \to \mathbb{R}^N$ is linear, then there exists an $N \times K$ matrix $\mathbf{A}$ such that $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^K$.

Now let's think about linear equations

The canonical problem: Find $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{b}$

In other words, find $\mathbf{x}$ such that $f(\mathbf{x}) = \mathbf{b}$

In other words, find $f^{-1}(\mathbf{b})$, the preimage of $\mathbf{b}$ under f

In other words, invert f at the point $\mathbf{b}$

Let's recall general problem of finding preimages

Two potential issues

- Existence fails
- Uniqueness fails

To build visual intuition, let's look at one-dimensional case

Let

- $f\colon [0,1] \to \mathbb{R}$ be some given function
- $b \in \mathbb{R}$

Want to find $x \in [0,1]$ such that $f(x) = b$

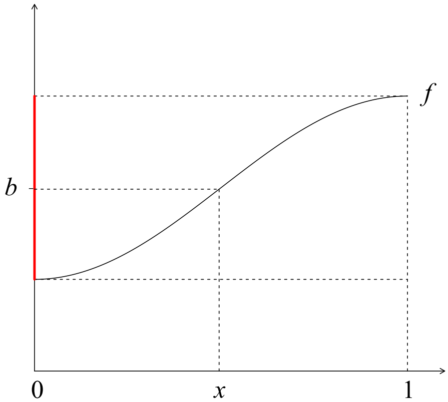


Figure: $x = f^{-1}(b)$ and red line is $\operatorname{rng}(f) := \{f(x) : x \in [0,1]\}$

Potential problems

- Existence fails: No $x \in [0, 1]$ such that $f(x) = b$
- Uniqueness fails: Many $x \in [0, 1]$ such that $f(x) = b$

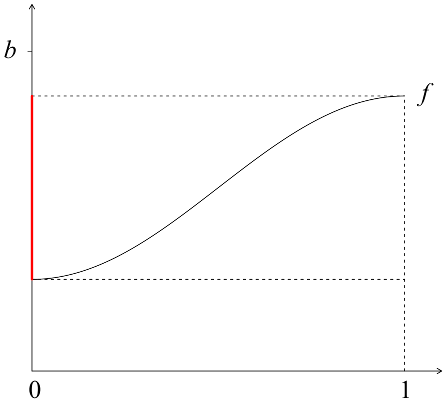


Figure: Existence fails because $b \notin \operatorname{rng}(f)$

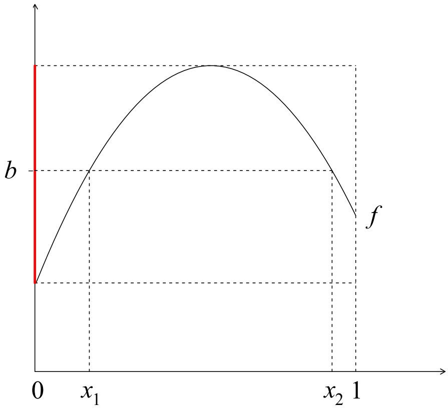


Figure: Uniqueness fails: $f(x_1) = f(x_2) = b$

Returning to matrix setting

- let $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$
- consider the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$

Solution exists if and only if $\mathbf{b}$ is in $\operatorname{rng}(f)$

This is more likely if the range of f is “large”

Standard measure of the “size” of the range is the rank of $\mathbf{A}$

Rank is related to linearly independence of its columns

Linear independence also connected to uniqueness of solutions

Let's learn more...

Linear Subspaces

Given $\mathbf{x}_1, \ldots, \mathbf{x}_K$ in $\mathbb{R}^N$ a **linear combination** is a vector

$$\mathbf{y} = \sum_{k=1}^{K} \alpha_k \mathbf{x}_k = \alpha_1 \mathbf{x}_1 + \cdots + \alpha_K \mathbf{x}_K$$

where $\alpha_1, \ldots, \alpha_K$ are scalars

Fact. Inner products of linear combinations satisfy

$$\left(\sum_{k=1}^{K} \alpha_k \mathbf{x}_k \right)' \left(\sum_{j=1}^{J} \beta_j \mathbf{y}_j \right) = \sum_{k=1}^{K} \sum_{j=1}^{J} \alpha_k \beta_j \mathbf{x}_k' \mathbf{y}_j$$

Linear Span of a Set

Set of all linear combinations of $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$ called the **span** of X, denoted by $\mathrm{span}(X)$:

$$\mathrm{span}(X) := \left\{ \text{all } \sum_{k=1}^{K} \alpha_k \mathbf{x}_k \text{ such that } (\alpha_1, \ldots, \alpha_K) \in \mathbb{R}^K \right\}$$

Example:

Let $X = \{\mathbf{1}\} \subset \mathbb{R}^2$, where $\mathbf{1} := (1, 1)$

The span of X is all vectors of the form (α, α) with $\alpha \in \mathbb{R}$

Constitutes a line in the plane

Example:

Consider the vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_N\} \subset \mathbb{R}^N$, where $\mathbf{e}_n$ has all zeros except for a 1 as the n-th element

$$\mathbf{e}_n := \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \leftarrow n\text{-th element}$$

Vectors $\mathbf{e}_1, \ldots, \mathbf{e}_N$ called the **canonical basis vectors** of $\mathbb{R}^N$

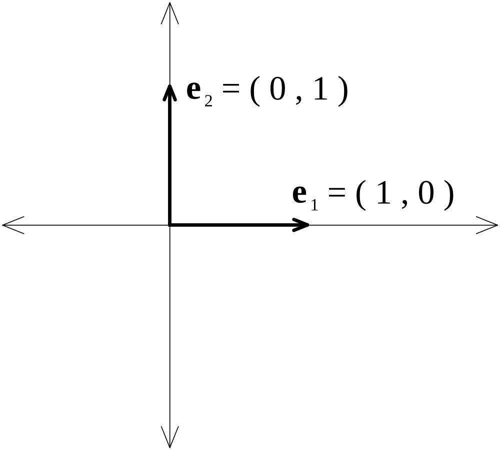


Figure: Canonical basis vectors in $\mathbb{R}^2$

Claim. The span of $\{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$ is equal to all of $\mathbb{R}^N$

Proof for $N = 2$:

Pick any $\mathbf{y} \in \mathbb{R}^2$

We have

$$\mathbf{y} := \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ y_1 \end{pmatrix}$$

$$= y_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = y_1 \mathbf{e}_1 + y_2 \mathbf{e}_2$$

Thus, $\mathbf{y} \in \mathrm{span}\{\mathbf{e}_1, \mathbf{e}_2\}$

Since $\mathbf{y}$ arbitrary, we have shown that $\mathrm{span}\{\mathbf{e}_1, \mathbf{e}_2\} = \mathbb{R}^2$

Example: Consider the set

$$P := \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1, x_2 \in \mathbb{R}\}$$

Graphically, P corresponds to the flat plane in $\mathbb{R}^3$, where height coordinate is always zero

For $\mathbf{e}_1 := (1, 0, 0)$ and $\mathbf{e}_2 := (0, 1, 0)$, we have $\operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2\} = P$

On one hand, given $\mathbf{y} = (y_1, y_2, 0) \in P$, we have $\mathbf{y} = y_1 \mathbf{e}_1 + y_2 \mathbf{e}_2$

In other words, $P \subset \operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2\}$

Conversely (check it) we have $\operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2\} \subset P$

Key feature of span: "Closed" under vector addition and scalar multiplication

Proof: If $\mathbf{y} \in \operatorname{span}(X)$ and $\mathbf{z} \in \operatorname{span}(X)$, then can write

$$\mathbf{y} = \sum_{k=1}^{K} \alpha_k \mathbf{x}_k \quad \text{and} \quad \mathbf{z} = \sum_{k=1}^{K} \beta_k \mathbf{x}_k$$

Follows that

$$\mathbf{y} + \mathbf{z} = \sum_{k=1}^{K} (\alpha_k + \beta_k) \mathbf{x}_k \in \operatorname{span}(X)$$

$$\delta \mathbf{y} = \sum_{k=1}^{K} (\delta \alpha_k) \mathbf{x}_k \in \operatorname{span}(X)$$

In general: If $S \subset \mathbb{R}^N$ closed under vector addition and scalar multiplication, then S called a linear subspace of $\mathbb{R}^N$

Formally: Nonempty $S \subset \mathbb{R}^N$ called a **linear subspace** if

$$\alpha \mathbf{x} + \beta \mathbf{y} \in S \quad \text{whenever} \quad \mathbf{x}, \mathbf{y} \in S \text{ and } \alpha, \beta \in \mathbb{R}$$

Examples of linear subspaces:

- The span of a collection X of vectors in $\mathbb{R}^N$
- $\mathbb{R}^N$ and $\{\mathbf{0}\}$ in $\mathbb{R}^N$
- Lines through origin in $\mathbb{R}^2$ and $\mathbb{R}^3$
- Planes through origin in $\mathbb{R}^3$

Exercise: Let S be a linear subspace of $\mathbb{R}^N$. Show that

1. $\mathbf{0} \in S$
2. If $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_K\} \subset S$, then $\operatorname{span}(X) \subset S$

Linear Independence

Example: Recall that $\text{span}\{\mathbf{e}_1, \mathbf{e}_2\}$ = all of $\mathbb{R}^2$

Removing either element of $\text{span}\{\mathbf{e}_1, \mathbf{e}_2\}$ reduces the span to a line

Now consider pair $\mathbf{e}_1$ and $\mathbf{x} := -2\mathbf{e}_1 = (-2, 0)$

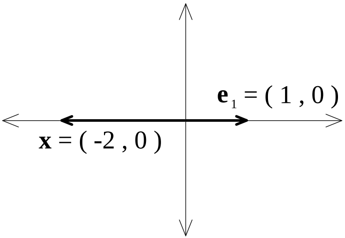


Figure: The vectors $\mathbf{e}_1$ and $\mathbf{x}$

Suppose $\mathbf{y} \in \mathrm{span}\{\mathbf{e}_1, \mathbf{x}\}$

Then, for some α_1 and α_2, we have $\mathbf{y} = \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{x}$

$$\therefore \quad \mathbf{y} = \alpha_1 \mathbf{e}_1 + \alpha_2(-2)\mathbf{e}_1 = (\alpha_1 - 2\alpha_2)\mathbf{e}_1$$

$$\therefore \quad \mathbf{y} \in \mathrm{span}\{\mathbf{e}_1\}$$

Follows that $\mathrm{span}\{\mathbf{e}_1, \mathbf{x}\} = \mathrm{span}\{\mathbf{e}_1\}$

Conclusion: We can kick $\mathbf{x}$ out without reducing the span

The formal concept:

$X := \{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$ called **linearly dependent** if one can be removed without changing span

X called **linearly independent** if not linearly dependent

Suppose $\{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$ linearly dependent

Suppose in particular that $\mathrm{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_K\} = \mathrm{span}\{\mathbf{x}_2, \ldots, \mathbf{x}_K\}$

Since $\mathbf{x}_1 \in \mathrm{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$, this equality implies that

$$\mathbf{x}_1 \in \mathrm{span}\{\mathbf{x}_2, \ldots, \mathbf{x}_K\}$$

Hence, there exist constants $\alpha_2, \ldots, \alpha_K$ with

$$\mathbf{x}_1 = \alpha_2 \mathbf{x}_2 + \cdots \alpha_K \mathbf{x}_K$$

In other words, $\mathbf{x}_1$ can be expressed as linear combination of other elements in X

Leads to second (equivalent) definition: X called linearly dependent if at least one vector can be written as a linear combination of the others

All of following statements are equivalent:

1. $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$ is linearly independent
2. If X_0 is a proper subset of X, then $\mathrm{span}(X_0)$ is a proper subset of $\mathrm{span}(X)$
3. No vector in X can be written as linear combination of others
4. If $\sum_{k=1}^{K} \alpha_k \mathbf{x}_k = \mathbf{0}$, then $\alpha_1 = \alpha_2 = \cdots = \alpha_K = 0$

Lemma. The set of canonical basis vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$ is linearly independent

Proof: If $\sum_{k=1}^{K} \alpha_k \mathbf{e}_k = \mathbf{0}$, then

$$(\alpha_1, \ldots, \alpha_K) = \sum_{k=1}^{K} \alpha_k \mathbf{e}_k = \mathbf{0}$$

$$\therefore \quad \alpha_1 = \cdots = \alpha_K = 0$$

Hence $\{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$ linearly independent

Uniqueness

Let $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_K\} \subset \mathbb{R}^N$

Let $\mathbf{y} \in \mathbb{R}^N$

We know that if $\mathbf{y} \in \mathrm{span}(X)$, then exists representation

$$\mathbf{y} = \sum_{k=1}^{K} \alpha_k \mathbf{x}_k$$

But when is this representation unique?

Theorem. Let $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_K\} \subset \mathbb{R}^N$, and let $\mathbf{y} \in \operatorname{span}(X)$. If X is linearly independent, then there exists one and only one set of scalars $\alpha_1, \ldots, \alpha_K$ such that $\mathbf{y} = \sum_{k=1}^K \alpha_k \mathbf{x}_k$.

Proof: Since $\mathbf{y}$ is in the span of X, exists at least one such set of scalars. If there are two, then

$$\mathbf{y} = \sum_{k=1}^{K} \alpha_k \mathbf{x}_k = \sum_{k=1}^{K} \beta_k \mathbf{x}_k$$

$$\therefore \quad \sum_{k=1}^{K} (\alpha_k - \beta_k) \mathbf{x}_k = \mathbf{0}$$

$$\therefore \quad \alpha_k = \beta_k \quad \text{for all} \quad k$$

In other words, the representation is unique

Let S be a linear subspace of $\mathbb{R}^N$

$B := \{\mathbf{b}_1, \ldots, \mathbf{b}_K\}$ called a **basis** of S if

1. B is linearly independent
2. $\mathrm{span}(B) = S$

Note that every element of S can be represented uniquely in terms of basis vectors

If $\mathbf{y} \in S$, then exists one and only one set of scalars $\alpha_1, \ldots, \alpha_K$ such that

$$\mathbf{y} = \sum_{k=1}^{K} \alpha_k \mathbf{b}_k$$

(Follows from previous theorem.)

Canonical basis vectors form a basis of $\mathbb{R}^N$

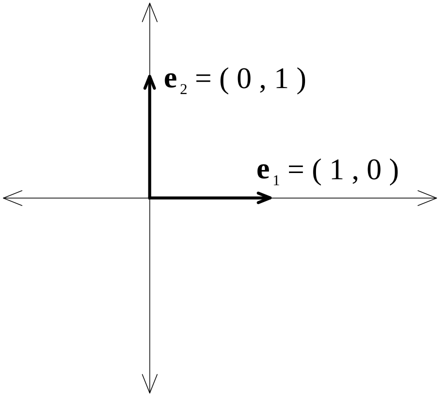


Figure: Canonical basis vectors in $\mathbb{R}^2$

Dimension

Theorem. If S is a linear subspace of $\mathbb{R}^N$, then every basis of S has the same number of elements

This common number is called the **dimension** of S

Example: $P := \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1, x_2 \in \mathbb{R}\}$ is two dimensional

- Spanned by $\{\mathbf{e}_1, \mathbf{e}_2\}$
- The set $\{\mathbf{e}_1, \mathbf{e}_2\}$ is linearly independent

Example: In $\mathbb{R}^3$, a line through the origin is one-dimensional, while a plane through the origin is two-dimensional

Fact. The only N-dimensional subspace of $\mathbb{R}^N$ is $\mathbb{R}^N$

Key question: How "big" is $\mathrm{span}(X)$?

Answer:

Theorem. Let X be a set of vectors with $\#X$ elements. The following statements are true:

1. $\dim(\mathrm{span}(X)) \leq \#X$
2. $\dim(\mathrm{span}(X)) = \#X$ if and only if X is linearly independent

Proof: See course notes

Rank

Let $\mathbf{A}$ be an $N \times K$ matrix and consider solving $f(\mathbf{x}) = \mathbf{A}\mathbf{x} = \mathbf{b}$

At least one such $\mathbf{x}$ will exist if $\mathbf{b}$ in

$$\mathrm{rng}(\mathbf{A}) := \mathrm{rng}(f) := \{\mathbf{y} \in \mathbb{R}^N : \mathbf{y} = \mathbf{A}\mathbf{x} \text{ for some } \mathbf{x} \in \mathbb{R}^K\}$$

Important: $\mathrm{rng}(\mathbf{A})$ is precisely the span of the columns of $\mathbf{A}$:

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

Sometimes called the **column space** of $\mathbf{A}$

Obviously a linear subspace of $\mathbb{R}^N$

To check $\mathbf{b} \in \mathrm{rng}(\mathbf{A})$ we want $\mathrm{rng}(\mathbf{A})$ to be "large"

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

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

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

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

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

Why do we say "full" column rank here?

Because $\operatorname{rng}(\mathbf{A})$ is the span of K vectors, and hence

$$\operatorname{rank}(\mathbf{A}) = \dim(\operatorname{rng}(\mathbf{A})) \leq K$$

$\mathbf{A}$ is said to have full col rank when this maximum is achieved

Therefore,

$\mathbf{A}$ is full column rank $\iff$ columns of $\mathbf{A}$ are linearly independent

The next characterization is also equivalent (why?)

Fact. $\mathbf{A}$ is of full column rank iff $\mathbf{A}\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$

Let's look again at solving $\mathbf{A}\mathbf{x} = \mathbf{b}$ for $\mathbf{x}$

There is a strong connection between existence and $\mathrm{rank}(\mathbf{A})$

For existence of a solution we need $\mathbf{b} \in \mathrm{rng}(\mathbf{A})$

The range will be large (in dimension) when $\mathbf{A}$ is full column rank

Good news: Rank is also closely connected to problem of uniqueness

Fact. If $\mathbf{A}$ has full column rank, then system $\mathbf{A}\mathbf{x} = \mathbf{b}$ has at most one solution

Important exercise: Why is this true?!

Square Matrices

Let $N = K$, so that $\mathbf{A}$ is a square $N \times N$ matrix

The perfect scenario:

Suppose that $\mathbf{A}$ is full column rank

Then (why?!) we must have $\operatorname{rng}(\mathbf{A}) = \mathbb{R}^N$

Follows immediately that, for any $\mathbf{b} \in \mathbb{R}^N$, the system $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution

Moreover, we know that the solution is unique

Since this problem is so important, there are several different ways of describing it

Square matrix $\mathbf{A}$ called **invertible** if there exists a $\mathbf{B}$ such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$

In this case, $\mathbf{B}$ called the **inverse** of $\mathbf{A}$, and written as $\mathbf{A}^{-1}$

Invertibility of $\mathbf{A}$ is equivalent to the existence of a unique solution to the system $\mathbf{Ax} = \mathbf{b}$ for all $\mathbf{b} \in \mathbb{R}^N$

The solution is given by

$$\mathbf{x}_b = \mathbf{A}^{-1}\mathbf{b} \tag{1}$$

Determinant

To each square matrix $\mathbf{A}$, can associate a unique number $\det(\mathbf{A})$ called the **determinant** of $\mathbf{A}$

Definition omitted—it won't come up

If $\det(\mathbf{A}) = 0$, then $\mathbf{A}$ is called **singular**

Otherwise $\mathbf{A}$ is called **nonsingular**

Fact. Nonsingularity is also equivalent to invertibility

Summary of invertibility for $N \times N$ matrix $\mathbf{A}$

Theorem. The following are equivalent:

1. $\mathbf{A}$ is of full column rank
2. The columns of $\mathbf{A}$ are linearly independent
3. $\mathbf{A}$ is invertible
4. $\mathbf{A}$ is nonsingular
5. $\mathbf{A}\mathbf{x} = \mathbf{b}$ has unique solution $\mathbf{A}^{-1}\mathbf{b}$ for each $\mathbf{b} \in \mathbb{R}^N$

Useful results about the inverse:

Fact. If $\mathbf{A}$ and $\mathbf{B}$ are invertible and $\alpha \neq 0$, then

1. $\det(\mathbf{A}^{-1}) = (\det(\mathbf{A}))^{-1}$
2. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
3. $(\alpha\mathbf{A})^{-1} = \alpha^{-1}\mathbf{A}^{-1}$
4. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

Other Properties of Matrices

The **transpose** of $N \times K$ matrix $\mathbf{A}$ is a $K \times N$ matrix $\mathbf{A}'$ such that the n-th column of $\mathbf{A}'$ is the n-th row of $\mathbf{A}$

For conformable matrices $\mathbf{A}$ and $\mathbf{B}$

1. $(\mathbf{A}\mathbf{B})' = \mathbf{B}'\mathbf{A}'$
2. $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$
3. $(c\mathbf{A})' = c\mathbf{A}'$ for any constant c.

For square matrix $\mathbf{A}$

1. $\det(\mathbf{A}') = \det(\mathbf{A})$
2. $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$

Trace of square matrix is sum of elements on principal diagonal

Transposition does not alter trace: $\text{trace}(\mathbf{A}) = \text{trace}(\mathbf{A}')$

For conformable matrices, any scalars:

- $\text{trace}(\alpha\mathbf{A} + \beta\mathbf{B}) = \alpha\,\text{trace}(\mathbf{A}) + \beta\,\text{trace}(\mathbf{B})$
- $\text{trace}(\mathbf{AB}) = \text{trace}(\mathbf{BA})$

Square matrix $\mathbf{A}$ called **idempotent** if $\mathbf{AA} = \mathbf{A}$

Fact. If $\mathbf{A}$ idempotent, then $\text{rank}(\mathbf{A}) = \text{trace}(\mathbf{A})$

For square symmetric $N \times N$ matrix $\mathbf{A}$ and $N \times 1$ vector $\mathbf{x}$,

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{j=1}^{N} \sum_{i=1}^{N} a_{ij} x_i x_j$$

called a **quadratic form** in $\mathbf{x}$

$\mathbf{A}$ called

- **nonnegative definite** if $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^N$
- **positive definite** if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \in \mathbb{R}^N$ with $\mathbf{x} \neq \mathbf{0}$

If $\mathbf{A}$ is nonnegative (positive) def, each a_{nn} nonnegative (positive)

Fact. If $\mathbf{A}$ is positive definite, then $\mathbf{A}$ is full column rank and $\det(\mathbf{A}) > 0$

Random Vectors and Matrices

A **random vector** $\mathbf{x}$ is a sequence of RVs $(x_1, \ldots, x_K)$

Distribution of $\mathbf{x}$ is the joint distribution F of $x_1, \ldots, x_K$:

$$\begin{aligned} F(\mathbf{s}) &:=: F(s_1, \ldots, s_K) \\ &:= \mathbb{P}\{x_1 \leq s_1, \ldots, x_K \leq s_K\} :=: \mathbb{P}\{\mathbf{x} \leq \mathbf{s}\} \end{aligned}$$

A **random $N \times K$ matrix** is an $N \times K$ array of random variables

We say that $f \colon \mathbb{R}^K \to \mathbb{R}$ is the density of $\mathbf{x}$ if

$$\int_B f(\mathbf{s})\, d\mathbf{s} = \mathbb{P}\{\mathbf{x} \in B\} \quad \text{for all} \quad B \subset \mathbb{R}^K \tag{2}$$

Random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_N$ called **independent** if, given any $\mathbf{s}_1, \ldots, \mathbf{s}_N$, we have

$$\mathbb{P}\{\mathbf{x}_1 \leq \mathbf{s}_1, \ldots, \mathbf{x}_N \leq \mathbf{s}_N\} = \mathbb{P}\{\mathbf{x}_1 \leq \mathbf{s}_1\} \times \cdots \times \mathbb{P}\{\mathbf{x}_N \leq \mathbf{s}_N\}$$

Fact. If $\mathbf{x}$ and $\mathbf{y}$ are independent and g and f are any functions, then $f(\mathbf{x})$ and $g(\mathbf{y})$ are also independent.

For random matrix $\mathbf{X}$, expectation is

$$\mathbb{E}[\mathbf{X}] := \begin{pmatrix} \mathbb{E}[x_{11}] & \mathbb{E}[x_{12}] & \cdots & \mathbb{E}[x_{1K}] \\ \mathbb{E}[x_{21}] & \mathbb{E}[x_{22}] & \cdots & \mathbb{E}[x_{2K}] \\ \vdots & \vdots & & \vdots \\ \mathbb{E}[x_{N1}] & \mathbb{E}[x_{N2}] & \cdots & \mathbb{E}[x_{NK}] \end{pmatrix}$$

Expectation of (vectors and) matrices also linear:

If $\mathbf{X}$ and $\mathbf{Y}$ are random and $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are conformable constant matrices, then

$$\mathbb{E}[\mathbf{A} + \mathbf{B}\mathbf{X} + \mathbf{C}\mathbf{Y}] = \mathbf{A} + \mathbf{B}\mathbb{E}[\mathbf{X}] + \mathbf{C}\mathbb{E}[\mathbf{Y}]$$

The **covariance** between random $N \times 1$ vectors $\mathbf{x}$ and $\mathbf{y}$ is

$$\operatorname{cov}[\mathbf{x}, \mathbf{y}] := \mathbb{E}\left[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{y} - \mathbb{E}[\mathbf{y}])'\right]$$

The **variance-covariance matrix** of $\mathbf{x}$

$$\operatorname{var}[\mathbf{x}] := \operatorname{cov}(\mathbf{x}, \mathbf{x}) = \mathbb{E}\left[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])'\right]$$

Expanding this out you will see that:

- The j, k-th term is the scalar covariance between x_j and x_k
- The principle diagonal contains the variance of each x_n

Exercise: Following alternative expressions are valid

- $\mathrm{cov}[\mathbf{x}, \mathbf{y}] = \mathbb{E}[\mathbf{x}\mathbf{y}'] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}]'$
- $\mathrm{var}[\mathbf{x}] = \mathbb{E}[\mathbf{x}\mathbf{x}'] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{x}]'$

Fact. For any random vector $\mathbf{x}$, the variance-covariance matrix $\mathrm{var}[\mathbf{x}]$ is square, symmetric and nonnegative definite

Fact. For random vector $\mathbf{x}$, constant conformable matrix $\mathbf{A}$ and constant conformable vector $\mathbf{a}$, we have

$$\mathrm{var}[\mathbf{a} + \mathbf{A}\mathbf{x}] = \mathbf{A}\,\mathrm{var}[\mathbf{x}]\mathbf{A}'$$

Multivariate normal density on $\mathbb{R}^N$ is

$$p(\mathbf{s}) = (2\pi)^{-N/2} \det(\mathbf{\Sigma})^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{s}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{s}-\boldsymbol{\mu})\right\}$$

- $\boldsymbol{\mu}$ is any $N \times 1$ vector
- $\mathbf{\Sigma}$ is a symmetric, positive definite $N \times N$ matrix

In symbols, we represent this distribution by $\mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$

Fact. If $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$, then

$$\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu} \quad \text{and} \quad \operatorname{var}[\mathbf{x}] = \mathbf{\Sigma}$$

We say that $\mathbf{x}$ is **standard normal** if $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{\Sigma} = \mathbf{I}$

Fact. $\mathbf{x}$ is normally distributed if and only if $\mathbf{a}'\mathbf{x}$ is normally distributed in $\mathbb{R}$ for every constant vector $\mathbf{a}$

Fact. If $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{a} + \mathbf{A}\mathbf{x} \sim \mathcal{N}(\mathbf{a} + \mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.

Fact. Normally distributed random variables are independent if and only if they are uncorrelated

Fact. If $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \sim \chi^2(k)$, where $k :=$ length of $\mathbf{x}$

Fact. If $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\mathbf{A}$ is idempotent and symmetric with $\operatorname{rank}(\mathbf{A}) = j$, then $\mathbf{x}'\mathbf{A}\mathbf{x} \sim \chi^2(j)$

Orthogonal Vectors and Projections

Let $\mathbf{x}$, $\mathbf{z}$ be vectors in $\mathbb{R}^N$

If $\mathbf{x}'\mathbf{z} = 0$, then $\mathbf{x}$ and $\mathbf{z}$ said to be **orthogonal**

Write $\mathbf{x} \perp \mathbf{z}$

In $\mathbb{R}^2$, orthogonal means perpendicular

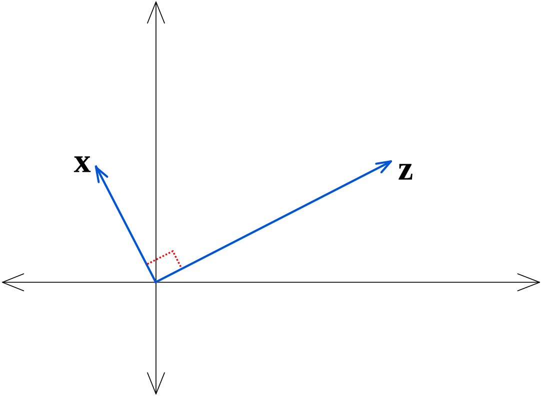


Figure: $\mathbf{x} \perp \mathbf{z}$

Let S be a linear subspace

We say that $\mathbf{x}$ **is orthogonal to** S if $\mathbf{x} \perp \mathbf{z}$ for all $\mathbf{z} \in S$

Write $\mathbf{x} \perp S$

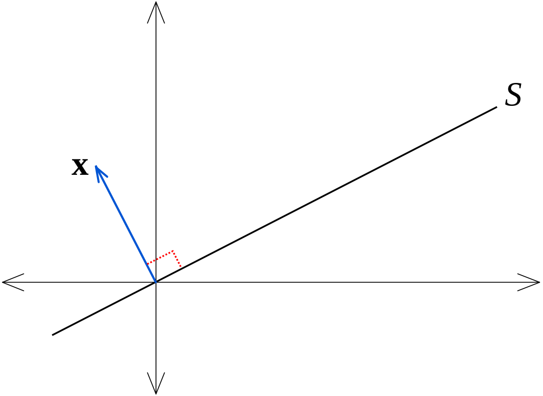


Figure: $\mathbf{x} \perp S$

Theorem. (Pythagorean Law) Let

- $\mathbf{x}_1, \ldots, \mathbf{x}_K \in \mathbb{R}^N$
- $\mathbf{x}_i \perp \mathbf{x}_j$ whenever $i \neq j$

Then

$$\|\mathbf{x}_1 + \cdots + \mathbf{x}_K\|^2 = \|\mathbf{x}_1\|^2 + \cdots + \|\mathbf{x}_K\|^2$$

Proof: Exercise.

Hint: Go back to the expression for inner products of linear combinations

Orthogonal Projections

Problem:

Given $\mathbf{y} \in \mathbb{R}^N$ and subspace S, find closest element of S to $\mathbf{y}$

Formally: Solve for

$$\hat{\mathbf{y}} := \operatorname*{argmin}_{\mathbf{z} \in S} \|\mathbf{y} - \mathbf{z}\|$$

Existence, uniqueness of solution not immediately obvious

Orthogonal projection theorem: $\hat{\mathbf{y}}$ always exists, unique

Also provides a useful characterization

Let $\mathbf{y} \in \mathbb{R}^N$ and let S be a linear subspace of $\mathbb{R}^N$

Theorem. (OPT Mark I) There exists a unique solution to

$$\hat{\mathbf{y}} := \operatorname*{argmin}_{\mathbf{z} \in S} \|\mathbf{y} - \mathbf{z}\|$$

The solution $\hat{\mathbf{y}}$ is the unique vector in $\mathbb{R}^N$ such that

1. $\hat{\mathbf{y}} \in S$
2. $\mathbf{y} - \hat{\mathbf{y}} \perp S$

Vector $\hat{\mathbf{y}}$ called the **orthogonal projection of $\mathbf{y}$ onto S**

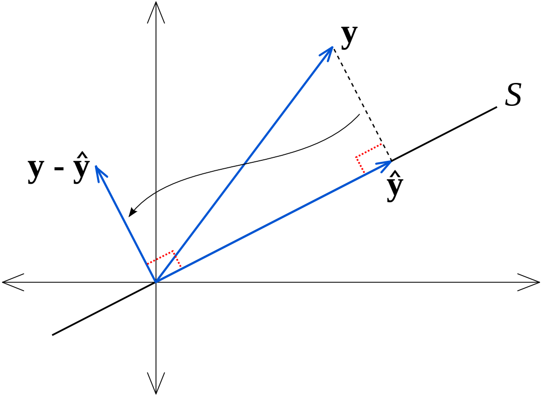


Figure: Orthogonal projection

Holding subspace S fixed, we have a functional relationship

$$\mathbf{y} \;\mapsto\; \text{its orthogonal projection } \hat{\mathbf{y}} \in S$$

A well-defined function from $\mathbb{R}^N$ to $\mathbb{R}^N$

The function is typically denoted by $\mathbf{P}$

- $\mathbf{P}(\mathbf{y})$ or $\mathbf{P}\mathbf{y}$ represents $\hat{\mathbf{y}}$

$\mathbf{P}$ is called the **orthogonal projection mapping onto S**

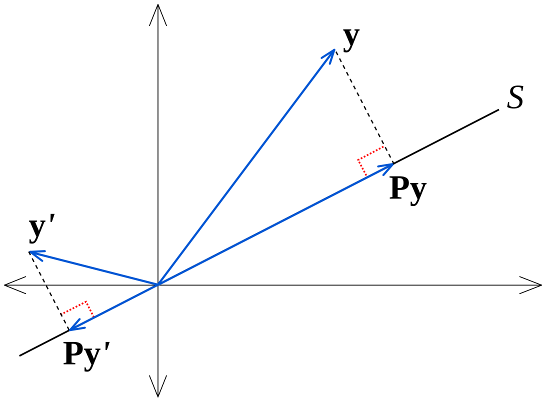


Figure: Orthogonal projection under $\mathbf{P}$

Fact. If S is any linear subspace of $\mathbb{R}^N$ and $\mathbf{P}$ is the orthogonal projection mapping onto S, then $\mathbf{P}$ is a linear function

Exercise: Prove it.

Let S be any linear subspace, and let $\mathbf{P}$ be the orthogonal projection mapping onto S

Theorem. (OPT Mark II) For any $\mathbf{y} \in \mathbb{R}^N$, we have

1. $\mathbf{P}\mathbf{y} \in S$,
2. $\mathbf{y} - \mathbf{P}\mathbf{y} \perp S$,
3. $\|\mathbf{y}\|^2 = \|\mathbf{P}\mathbf{y}\|^2 + \|\mathbf{y} - \mathbf{P}\mathbf{y}\|^2$,
4. $\|\mathbf{P}\mathbf{y}\| \leq \|\mathbf{y}\|$, and
5. $\mathbf{P}\mathbf{y} = \mathbf{y}$ if and only if $\mathbf{y} \in S$.

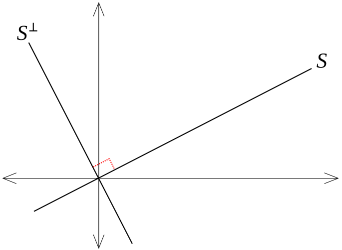
Proof of 3:

- Observe $\mathbf{y} = \mathbf{P}\mathbf{y} + \mathbf{y} - \mathbf{P}\mathbf{y}$, apply ?

Orthogonal Complement

Given linear subspace S, the **orthogonal complement** of S is

$$S^{\perp} := \{\mathbf{x} \in \mathbb{R}^N : \mathbf{x} \perp S\}$$



Fact. Orthogonal complement $S^\perp$ is always a linear subspace

Proof: Given $\mathbf{x}, \mathbf{y} \in S^\perp$ and $\alpha, \beta \in \mathbb{R}$, claim that $\alpha \mathbf{x} + \beta \mathbf{y} \in S^\perp$

True, because if $\mathbf{z} \in S$, then

$$(\alpha \mathbf{x} + \beta \mathbf{y})' \mathbf{z} = \alpha \mathbf{x}' \mathbf{z} + \beta \mathbf{y}' \mathbf{z} = \alpha \times 0 + \beta \times 0 = 0$$

$$\therefore \quad \alpha \mathbf{x} + \beta \mathbf{y} \in S^\perp$$

Fact. For any $S \subset \mathbb{R}^N$, we have $S \cap S^\perp = \{\mathbf{0}\}$

Orthogonal projection theorem, take 3

Our interest was in projecting $\mathbf{y}$ onto S

Can also project $\mathbf{y}$ onto $S^{\perp}$

Notation 1:

- $\mathbf{P}$ denotes projection onto S
- $\mathbf{M}$ denotes projection onto $S^{\perp}$

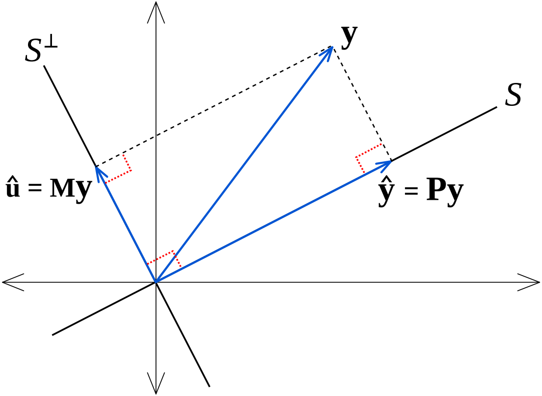


Figure: Orthogonal projection

Let S be a linear subspace of $\mathbb{R}^N$

Theorem. (OPT Mark III) If $\mathbf{P}$ is the orthogonal projection onto S and $\mathbf{M}$ is the orthogonal projection onto $S^\perp$, then $\mathbf{Py}$ and $\mathbf{My}$ are orthogonal, and

$$\mathbf{y} = \mathbf{Py} + \mathbf{My}$$

Fact. If $S_1 \subset S_2$, then, for any $\mathbf{y} \in \mathbb{R}^N$,

$$\mathbf{P}_1 \mathbf{P}_2 \mathbf{y} = \mathbf{P}_2 \mathbf{P}_1 \mathbf{y} = \mathbf{P}_1 \mathbf{y}$$

Fact. $\mathbf{P}\mathbf{y} = \mathbf{0}$ iff $\mathbf{y} \in S^{\perp}$, and $\mathbf{M}\mathbf{y} = \mathbf{0}$ iff $\mathbf{y} \in S$

Content of The Midterm Exam

The intersection of

- Chapters 3–6 of the course notes (inclusive)
- Material covered in the lectures

Hints for the Midterm

Know your definitions!!

If you don't know the answer, at least state any relevant definitions

To maximize your marks, be careful and rigorous

But don't be too hung up on being rigorous—you'll be under time pressure!

Focus on getting the answer as best you can. Come back and fill in more details if you have time.