

Academic Writing

Lecture 1

2 October 2026

Course Information

Goal

Good theses

Organisation

8 lectures and exercise classes

(no classes during the weeks of 5–9 October, 26–30 October, and 9–13 November)

Assignments

Reading/writing a text

Exercise classes

We will discuss the assignments together in class.

Exercise classes

A final written assignment, due by the end of the semester.

Why You Should Write Well

Limited attention

The reader's attention is a resource.

Attention spent hunting for definitions, dealing with typos, or navigating a disorganised text is no longer available for understanding.

Why You Should Write Well

Limited attention

The reader's attention is a resource.

Attention spent hunting for definitions, dealing with typos, or navigating a disorganised text is no longer available for understanding.

The cost of writing and the cost of reading

Increase the effort of the author (who writes once) to save the effort of many readers (who read repeatedly).

Why You Should Write Well

Limited attention

The reader's attention is a resource.

Attention spent hunting for definitions, dealing with typos, or navigating a disorganised text is no longer available for understanding.

The cost of writing and the cost of reading

Increase the effort of the author (who writes once) to save the effort of many readers (who read repeatedly).

A word of caution

Too much explanation is not good either. (Attention is a resource.)

Before You Start Writing

Before You Start Writing: Read

- What is already known?
- What is not known?
- What is important?
- What interests people?
- Is the topic new?
- Will the topic remain relevant?

What to look for

- People and groups working on the topic
- Browse the references in papers
- Look for other results

Before You Start Writing

Before You Start Writing: Read

- What is already known?
- What is not known?
- What is important?
- What interests people?
- Is the topic new?
- Will the topic remain relevant?

What to look for

- People and groups working on the topic
- Browse the references in papers
- Look for other results

How to search

- Abstracts (this is where the results are)
- Introductions (this is where previous results and citations are)
- Reference lists in general
- Who has cited the papers (there are search engines for this)
- Ask AI

Purpose, Format $\Rightarrow$ Style

There are many ways to write the same text.

Key differences:

What is the format?

- Textbook
- Research paper
- Student paper
- Popular science article
- Presentation?

Purpose, Format $\Rightarrow$ Style

There are many ways to write the same text.

Key differences:

What is the format?

- Textbook
- Research paper
- Student paper
- Popular science article
- Presentation?

What is the purpose?

- Presenting results
- Demonstrating skills
- Being entertaining
- Providing useful course notes

Results

- New
- Known

Results

- New
- Known

What does this affect?

- Whether you need to demonstrate your skills
- Whether you need to convince the reader of correctness
- Whether you need to identify who contributed each part

Who is the intended reader?

- An average person
- Someone familiar with the field
- An expert
- ...

This is not the same as the actual reader: master's theses are read by experts, but should be accessible to a broad audience.

The Structure of a Paper

The bad: a stream of consciousness

The Structure of a Paper

The bad: a stream of consciousness

The ugly

(Definition, Lemma, Theorem) $^{\infty}$

The Structure of a Paper

The good

A paper is a story. (Read Chandler.)

The bad: a stream of consciousness

The ugly

(Definition, Lemma, Theorem) $^{\infty}$

Example

- ❶ **Problem:** find a shortest path in a graph with negative edge weights but no negative cycles.
- ❷ **Obstacle:** Dijkstra's algorithm does not work with negative edge weights.
- ❸ **Idea:** reweight the edges using a potential function.
- ❹ **Lemma:** reweighting preserves the ordering of path lengths.
- ❺ **Theorem:** after reweighting, Dijkstra's algorithm can be applied.

Example

- ➊ **Problem:** find a shortest path in a graph with negative edge weights but no negative cycles.
- ➋ **Obstacle:** Dijkstra's algorithm does not work with negative edge weights.
- ➌ **Idea:** reweight the edges using a potential function.
- ➍ **Lemma:** reweighting preserves the ordering of path lengths.
- ➎ **Theorem:** after reweighting, Dijkstra's algorithm can be applied.

Sequence: problem $\rightarrow$ obstacle $\rightarrow$ idea $\rightarrow$ tool $\rightarrow$ result.

Example

- ❶ **Problem:** the existing algorithm is too slow for large instances.
- ❷ **Hypothesis:** the expensive part can be approximated.
- ❸ **Method:** we introduce heuristic H.
- ❹ **Experiment:** we compare A, A+H, and a simple baseline.
- ❺ **Result:** H reduces running time, but reduces quality only on some instances.

Example

Weak

Definition 2.1 $\rightarrow$ Lemma 2.2 $\rightarrow$ Lemma 2.3 $\rightarrow$ Theorem 2.4.

Example

Weak

Definition 2.1 $\rightarrow$ Lemma 2.2 $\rightarrow$ Lemma 2.3 $\rightarrow$ Theorem 2.4.

A better narrative

“First, we define the decomposition P . Next, we show that the construction preserves property Q (Lemma 2.2). This property is crucial to the proof of the main result (Theorem 2.4).”

The Structure of a Paper

The narrative nature of the text

The structure of a paper

Abstract → Introduction → Definitions → Results → Conclusions

Abstract → Introduction → Methods, equipment description, evaluation methodology → Results → Analysis → Conclusions

is not arbitrary. The goal is to tell a coherent story about the results.

The Structure of a Paper

The narrative nature of the text

The structure of a paper

Abstract → Introduction → Definitions → Results → Conclusions

Abstract → Introduction → Methods, equipment description, evaluation methodology → Results → Analysis → Conclusions

is not arbitrary. The goal is to tell a coherent story about the results.

- At every point, the reader should know:
 - 1 What have we done?
 - 2 What are we doing now?
 - 3 What will we do next?

The Structure of a Paper

The narrative nature of the text

The structure of a paper

Abstract → Introduction → Definitions → Results → Conclusions

Abstract → Introduction → Methods, equipment description, evaluation methodology → Results → Analysis → Conclusions

is not arbitrary. The goal is to tell a coherent story about the results.

- At every point, the reader should know:

- ① What have we done?
- ② What are we doing now?
- ③ What will we do next?

This applies at a high level: what the chapters tell us.

At a low level: is this a definition or a proof?

Is this an assumption or a goal?

An interesting fact: Newton's original works are still being analysed (by enthusiasts).

The number of reconstructed proofs keeps increasing!

The Structure of a Paper

The narrative nature of the text

The structure of a paper

Abstract → Introduction → Definitions → Results → Conclusions

Abstract → Introduction → Methods, equipment description, evaluation methodology → Results → Analysis → Conclusions

is not arbitrary. The goal is to tell a coherent story about the results.

- At every point, the reader should know:

- ① What have we done?
- ② What are we doing now?
- ③ What will we do next?

This applies at a high level: what the chapters tell us.

At a low level: is this a definition or a proof?

Is this an assumption or a goal?

An interesting fact: Newton's original works are still being analysed (by enthusiasts).

The number of reconstructed proofs keeps increasing!

- **Writing order** ≠ **reading order**: the abstract and introduction are usually written last!

Outline

Before you start writing, make an outline.
A reasonably detailed one.

Outline

Before you start writing, make an outline.
A reasonably detailed one.

- The order
- What each chapter should convey
- What each definition and theorem should convey

Outline

Before you start writing, make an outline.
A reasonably detailed one.

- The order
- What each chapter should convey
- What each definition and theorem should convey

Changing the outline

The outline will change, and that is normal.

Writing and revising is a highly iterative process.

“Encapsulation”

The reader (and usually the author, too) cannot keep the entire paper in mind at once. You cannot write on the assumption that anything mentioned earlier can be freely used later.

“Encapsulation”

The reader (and usually the author, too) cannot keep the entire paper in mind at once. You cannot write on the assumption that anything mentioned earlier can be freely used later.

Two consequences:

- If something is to be used (much) later, give it a name: a lemma, invariant, property, table, numbered equation, list, ... Then refer to it explicitly.
 - “Hide” the complexity (encapsulation): a lemma with a clearly stated claim, with the difficult construction hidden inside.
- It is worth breaking things up to make proofs shorter.

“Encapsulation”

The reader (and usually the author, too) cannot keep the entire paper in mind at once. You cannot write on the assumption that anything mentioned earlier can be freely used later.

Two consequences:

- If something is to be used (much) later, give it a name: a lemma, invariant, property, table, numbered equation, list, ... Then refer to it explicitly.
- “Hide” the complexity (encapsulation): a lemma with a clearly stated claim, with the difficult construction hidden inside.
It is worth breaking things up to make proofs shorter.
- **Lemma (internal use)**: hides difficult technical details. You can read it, understand it, and forget its inner workings.
- **Theorem (external use)**: the main result. Its statement should not require familiarity with local, highly specific notation from the proof.

Example

AVL invariant

The heights of the subtrees differ by at most 1.

Example

AVL invariant

The heights of the subtrees differ by at most 1.

Lemma

The AVL invariant can be restored in $\mathcal{O}(\log n)$ time for operations ...

Example

AVL invariant

The heights of the subtrees differ by at most 1.

Lemma

The AVL invariant can be restored in $\mathcal{O}(\log n)$ time for operations ...

Theorem

AVL trees have height $\mathcal{O}(\log n)$. Operations ... take $\mathcal{O}(\log n)$ time.

Example

Lemma

“For the structure $\mathcal{H}_k$, we have $\text{preproc}(\mathcal{H}_k) \leq O(n)$, $\text{query}(\mathcal{H}_k) \leq O(1)$.”

Example

Lemma

“For the structure $\mathcal{H}_k$, we have $\text{preproc}(\mathcal{H}_k) \leq O(n)$, $\text{query}(\mathcal{H}_k) \leq O(1)$.”

Theorem

For every binary tree T with n vertices, after $O(n)$ -time preprocessing, an ancestor query takes $O(1)$ time.

Example: Stating a Theorem

Lemma, weak: “For the structure $\mathcal{H}_k$ defined in the proof of Lemma 2, we have $\text{cost}(\mathcal{H}_k) \leq O(n)$.”

Lemma, so-so: “For the structure $\mathcal{H}_k$, we have $\text{cost}(\mathcal{H}_k) \leq O(n)$.”

Lemma, good: “For the structure $\mathcal{H}_k$, we have $\text{cost}(\mathcal{H}_k) \leq O(|\mathcal{H}_k|)$.”

Theorem: “For every binary tree with n vertices, the construction time is $O(n)$.”

Example: Stating a Theorem

Lemma, weak: “For the structure $\mathcal{H}_k$ defined in the proof of Lemma 2, we have $\text{cost}(\mathcal{H}_k) \leq O(n)$.”

Lemma, so-so: “For the structure $\mathcal{H}_k$, we have $\text{cost}(\mathcal{H}_k) \leq O(n)$.”

Lemma, good: “For the structure $\mathcal{H}_k$, we have $\text{cost}(\mathcal{H}_k) \leq O(|\mathcal{H}_k|)$.”

Theorem: “For every binary tree with n vertices, the construction time is $O(n)$.”

Example: Stating Conclusions

Internal statement, weak: “Algorithm A achieved a score 0.1 higher than algorithm B on the corpus used.”

Internal statement, better: “On the Pizza&Chilli corpus, the RePair algorithm achieved a compression result 5–10% better than the LZ78 algorithm.”

Conclusions: “The RePair algorithm proposed in this paper, based on replacing the most frequent repetitions and entropy coding, achieved a compressed data size 5–10% better on the standard Pizza&Chilli corpus than the standard LZ78 algorithm, which is based on ... The difference is due to ...”

[D. Knuth]: Explain Twice

Intuition before formalism

Explain every complicated concept **twice**:
first intuitively and informally, and only then rigorously and formally.

Intuition before formalism

Explain every complicated concept **twice**:
first intuitively and informally, and only then rigorously and formally.

1. **Intuition:** “The algorithm splits the sequence in half, sorts the halves recursively, and then merges them while preserving order.”
2. **Formalism:** “Let $A = (a_1, \dots, a_n)$. We define the partition $P(A) = (A_1, A_2)$, where $A_1 = (a_1, \dots, a_{\lfloor n/2 \rfloor}) \dots$ ”

A description can also serve as a proof: the principle behind rotations in AVL trees.

Intuition before formalism

Explain every complicated concept **twice**:
first intuitively and informally, and only then rigorously and formally.

1. **Intuition:** “The algorithm splits the sequence in half, sorts the halves recursively, and then merges them while preserving order.”
2. **Formalism:** “Let $A = (a_1, \dots, a_n)$. We define the partition $P(A) = (A_1, A_2)$, where $A_1 = (a_1, \dots, a_{\lfloor n/2 \rfloor}) \dots$ ”

A description can also serve as a proof: the principle behind rotations in AVL trees.

- Pure formalism alone is difficult to follow.
- Intuition is robust to minor calculation errors and typos in the proof.
- Good intuition can suffice as a proof.

Intuition in Heuristics

Intuition in Heuristics

For a given sequence of input features $X \in \mathbb{R}^{N \times d}$, where N is the sequence length and d is the hidden dimension, the attention mechanism defines three matrices:

- queries (Q),
- keys (K),
- values (V).

These are obtained through linear transformations with learnable weights W^Q, W^K, W^V :

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V \quad (1)$$

The output of the attention operation is computed using the equation:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \quad (2)$$

where d_k is the key dimension, used as a scaling factor to prevent the *softmax* function from entering regions with near-zero gradients.

The Purpose of Proofs and Describing Algorithms

What is the purpose of a proof?

The Purpose of Proofs and Describing Algorithms

What is the purpose of a proof?

An academic text is not input to a theorem prover. The purpose of a proof is to **convince** the reader and help them understand **why** the result is true.

A common question: “Does the proof have to be formal?”

The same applies to heuristic approaches.

The Purpose of Proofs and Describing Algorithms

What is the purpose of a proof?

An academic text is not input to a theorem prover. The purpose of a proof is to **convince** the reader and help them understand **why** the result is true.

A common question: “Does the proof have to be formal?”

The same applies to heuristic approaches.

- **Presenting algorithms:**

- Intuition
- Precision
- Correctness
- Source code alone is difficult for the reader to follow.
- Natural language alone is imprecise.
- A natural-language description (high level) + clear pseudocode + proof/intuition.

$f \leftarrow 1, g \leftarrow 0$

for $m \leftarrow 0, \dots, 2t - 1$ **do**

$c \leftarrow \deg(f) - 1$

$r \leftarrow m - c - 1$

5:

$\mu \leftarrow \vec{e} \cdot \text{shift}(f, r)$

if $\mu \neq 0$ **then**

if $r \leq c$ **then**

10:

$f' \leftarrow f - \mu \text{shift}(g, c - r)$

$g' \leftarrow g$

else

$f' \leftarrow \text{shift}(f, r - c) - \mu g$

$g' \leftarrow \mu^{-1} f$

15:

return $f, g \leftarrow f', g'$

▷ t is not needed; an upper bound suffices

▷ $r = m - \deg(f)$

▷ Invariants: $\text{Span}(f) \geq r$, $\text{Span}(g) = c$, $\text{disc}(g) = 1$

▷ Possible for $m \leq 2t - 1 \leq n - k - 1$

▷ If $\mu = 0$, then f, g are unchanged

▷ Here we have $\text{Span}(f) = r$.

▷ The update from Lemma 4

▷ The update from Lemma 3, multiplied by $-\mu^{-1}$

Two roles of citations

1. Pointing to related work, problems, and results (mainly in the introduction).
2. Clearly attributing authorship (crucial in the rest of the text).

- In a research paper, the default assumption is: no citation means that the result and proof are the author's own.
- Theses:
 - Often reproduce known facts. This must be stated explicitly.
 - A literal translation or copy of a definition, theorem, proof, pseudocode, example, ... requires clear acknowledgement.

Two roles of citations

1. Pointing to related work, problems, and results (mainly in the introduction).
2. Clearly attributing authorship (crucial in the rest of the text).

- In a research paper, the default assumption is: no citation means that the result and proof are the author's own.
- Theses:
 - Often reproduce known facts. This must be stated explicitly.
 - A literal translation or copy of a definition, theorem, proof, pseudocode, example, ... requires clear acknowledgement.

"The statement of Theorem 3.2 comes from [Kowalski, 2020], but we have simplified and corrected the proof presented below (Case 3.1)."

Precise citations:

- [5]
- ✓ [5, Theorem 3.1]
- [5, Theorem 3.1], the proof was given for the case $n = 2$, but also works for $n = 3$

Too vague

“The graph colouring theorem is known.”

Too vague

“The graph colouring theorem is known.”

Good

“It is well known that every graph with maximum degree Δ is $(\Delta + 1)$ -colourable; see, for example, [Theorem 3.1, 5].”

Too vague

“The graph colouring theorem is known.”

Good

“It is well known that every graph with maximum degree Δ is $(\Delta + 1)$ -colourable; see, for example, [Theorem 3.1, 5].”

Good

“We use Theorem 3.1 from [5], which states that every graph with maximum degree Δ is $(\Delta + 1)$ -colourable.”

Too vague

“The graph colouring theorem is known.”

Good

“It is well known that every graph with maximum degree Δ is $(\Delta + 1)$ -colourable; see, for example, [Theorem 3.1, 5].”

Good

“We use Theorem 3.1 from [5], which states that every graph with maximum degree Δ is $(\Delta + 1)$ -colourable.”

Good

“The graph colouring theorem [5, Theorem 3.1] is usually stated for finite graphs, but the proof readily extends to infinite graphs.”

Theorem (Nowak, Kowalski [1978])

...

Footnote

The result in Nowak, Kowalski [1978] is stated for one of the solutions, but the proof works with little modification for all minimal solutions.

Theorem (Nowak, Kowalski [1978])

...

Footnote

The result in Nowak, Kowalski [1978] is stated for one of the solutions, but the proof works with little modification for all minimal solutions.

- Precision: where exactly is this in the paper?
- When asked, Nowak is not sure whether it is true.
- The Nowak–Kowalski paper contains a serious error (and it is unclear whether the result is true at all).

When we cite, we take responsibility.