# Introduction

The course is split into two parts:

- Logic: syntax and semantics.
- Set theory: what does the universe of sets look like?

Course structure

(I) Propositional logic (logic)

(II) Well-orderings & ordinals (set theory)

(III) Posets & Zorn's lemma (set theory)

(IV) Predicate logic (logic)

(V) Set theory (set theory)

(VI) Cardinals (set theory)

Books:

1. Johnstone, *Notes on Logic & Set Theory*
2. Van Dalen, *Logic & Structure* (Chapter 4 and what 'goes next')
3. Hajnal & Hamburger, *Set Theory* (Chapters 2 and 6)
4. Forster, *Logic, Induction & Sets*

# 1 Propositional Logic

Let $P$ be a set of *primitive propositions*. Unless otherwise stated, $P = \{p_1, p_2, \ldots\}$. The *language* $L$ or $L(P)$ is defined inductively by

1. If $p \in P$, then $p \in L$
2. $\perp \in L$ ($\perp$ is read 'false')
3. If $p, q \in L$ then $(p \Rightarrow q) \in L$.

e.g $((p_1 \Rightarrow p_2) \Rightarrow (p_1 \Rightarrow p_3)), (p_4 \Rightarrow \perp), (\perp \Rightarrow \perp)$.

**Notes.**

1. Each proposition (member of $L$) is a finite string of symbols from language: $\vdash, \Rightarrow, \perp, p_1, p_2, \ldots$ (for clarity often omit outer brackets, use other types of bracket, etc).
2. '$L$ is defined inductively' means, more precisely, the following

- Put $L_1 = P \cup (\perp)$;
- Having defined $L_n$, put $L_{n+1} = L_n \cup \{(p \Rightarrow q) : p, q \in L_n\}$;
- Set $L = \bigcup_{n \geq 1} L_n$.

3. Every $p \in L$ is uniquely built up from steps 1,2 using 3. For example, $((p_1 \Rightarrow p_2) \Rightarrow (p_1 \Rightarrow p_3))$ can from $(p_1 \Rightarrow p_2)$ and $(p_1 \Rightarrow p_3)$.

We can now introduce $\neg p$ ('not $p$') as an abreviation for $(p \Rightarrow \perp)$; $p \vee q$ ('$p$ or $q$') as an abreviation for $(\neg p) \Rightarrow q$; $p \wedge q$ ('$p$ and $q$') as an abreviation for $\neg(p \Rightarrow (\neg q))$.

## 1.1 Semantic Implication

**Definition.** A *valuation* is a function $v : L \to \{0,1\}$ (thinking of 0 as 'False' and 1 as 'True') such that

(i) $v(\perp) = 0$

(ii) $v(p \Rightarrow q) = \begin{cases} 0 & \text{if } v(p) = 1,\ v(q) = 0 \\ 1 & \text{otherwise} \end{cases}$.

**Remark.** On $\{0,1\}$, could define a constant $\perp = 0$ and an operation $\Rightarrow$ by

$$(a \Rightarrow b) = \begin{cases} 0 & \text{if } a = 1, b = 0 \\ 1 & \text{otherwise} \end{cases}.$$

Then a valuation is precisely a mapping $L \to \{0,1\}$ that preserves ($\perp$ and $\Rightarrow$).

**Proposition 1.1.**

*(i) If $v, v'$ are valuations with $v(p) = v(p')$ for all $p \in P$, then $v = v'$.*

*(ii) For any function $w : P \to \{0,1\}$, there exists a valuation $v$ with $v(p) = w(p)$ for all $p \in P$.*

*Proof.*

(i) Have $v(p) = v'(p)$ for all $p \in L_1$. But if $v(p) = v'(p)$ and $v(q) = v'(q)$, then $v(p \Rightarrow q) = v'(p \Rightarrow q)$, so $v(p) = v'(p)$ for all $p \in L_2$. Continuing inductively we obtain $v(p) = v'(p)$ for all $p \in L_n$ for each $n$.

(ii) Set $v(p) = w(p)$ for all $p \in P$ and $v(\perp) = 0$ to obtain $v$ on $L_1$. Now put

$$v(p \Rightarrow q) = \begin{cases} 0 & v(p) = 1, v(q) = 0 \\ 1 & \text{otherwise} \end{cases}$$

to obtain $v$ on $L_2$, then induction.

$\square$

**Example.** Let $v$ be the valuation with $v(p_1) = v(p_3) = 1$, $v(p_n) = 0$ for all $n \neq 1, 3$. Then $v((p_1 \Rightarrow p_2) \Rightarrow p_3) = 0$.

**Definition.** A *tautology* is an element $t \in L$ such that $v(t) = 1$ for any valuation $v$. We write $\models t$.

**Examples.**

1. $p \Rightarrow (q \Rightarrow p)$

| $v(p)$ | $v(q)$ | $v(p \Rightarrow q)$ | $v(p \Rightarrow (q \Rightarrow p))$ |
|---|---|---|---|
| 0 | 0 | 1 | 1 |
| 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 1 |
| 1 | 1 | 1 | 1 |

So this is a tautology.

2. $(\neg\neg p) \Rightarrow p$, i.e $((p \Rightarrow \perp) \Rightarrow \perp) \Rightarrow p$ ('law of excluded middle')

| $v(p)$ | $v(p \Rightarrow \perp)$ | $v((p \Rightarrow \perp) \Rightarrow \perp)$ | $v(((p \Rightarrow \perp) \Rightarrow \perp) \Rightarrow p)$ |
|---|---|---|---|
| 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 1 |

3. $(p \Rightarrow (q \Rightarrow r)) \Rightarrow ((p \Rightarrow q) \Rightarrow (p \Rightarrow r))$ ("how implicatino chains"). Suppose this is not a tautology. Then we have a $v$ with $v(p \Rightarrow (q \Rightarrow r)) = 1$ and $v((p \Rightarrow q) \Rightarrow (q \Rightarrow r)) = 0$. Then $v(p \Rightarrow q) = 1$ and $v(p \Rightarrow r) = 0$. Hence $v(p) = 1$ and $v(r) = 0$, so $v(q) = 1$. Hence $v(p \Rightarrow (q \Rightarrow r)) = 0$, contradiction.

**Definition.** For $S \subseteq L$, $t \in L$, we say $S$ *entails* or *semantically implies* $t$, written $S \vDash t$ if every valuation with $v(s) = 1$ for all $s \in S$ has $v(t) = 1$.

**Example.** $\{p \Rightarrow q, q \Rightarrow r\}$ entails $p \Rightarrow r$. Indeed, suppose we have $v$ with $v(p \Rightarrow q)$, $v(q \Rightarrow r) = 1$ but $v(p \Rightarrow r)$. Then $v(p) = 1$, $v(r) = 0$. Hence $v(q) = 1$, contradicting $v(q \Rightarrow r) = 1$.

**Definition.** We say $v$ is a *model* of $S \subseteq L$ or $S$ is *true* in $v$, if $v(s) = 1$ for all $s \in S$. Thus $S$ entails $t$ means: every model of $S$ is also a model of $\{t\}$.

**Remark.** $\vDash t$ says $\emptyset \vDash t$.

## 1.2 Syntatic implication

For a notion of proof, we'll need axioms and deduction rules. As axioms, we'll take:

1. $p \Rightarrow (q \Rightarrow p)$ for all $p, q \in L$;
2. $[p \Rightarrow (q \Rightarrow r)] \Rightarrow [(p \Rightarrow q) \Rightarrow (p \Rightarrow r)]$ for all $p, q \in L$;
3. $(\neg\neg p) \Rightarrow p$ for all $p \in L$.

**Notes.**

1. Sometimes we call these 'axiom schemes' since each is actually a set of axioms.
2. Each of these are tautologies.

For deduction rules, we'll have only *modus ponens*: from each $p$ and $p \Rightarrow q$ we can deduce $q$.

**Definition.** For $S \subseteq L$, and $t \in S$, say $S$ *proves* or *syntactically implies* $t$, written $S \vdash t$ if there exists a sequence $t_1, \ldots, t_n$ in $L$ with $t_n = t$ such that every $t_i$ is either

(i) An axiom; or

(ii) A member of $S$; or

(iii) Such that there exist $j, k < i$ with $t_k \Rightarrow (t_j \Rightarrow t_n)$ (modus ponens).

Say $S$ consists of the *hypotheses* or *premises*, and $t$ the *conclusion*.

**Example.** $\{p \Rightarrow q, q \Rightarrow r\} \vdash p \Rightarrow r$:

1. $q \Rightarrow r$ (hypothesis)
2. $(q \Rightarrow r) \Rightarrow (p \Rightarrow (q \Rightarrow r))$ (axiom 1)
3. $p \Rightarrow (q \Rightarrow r)$ (modus ponens' on 2,3)
4. $[p \Rightarrow (q \Rightarrow r)] \Rightarrow [(p \Rightarrow q) \Rightarrow (p \Rightarrow r)]$ (axiom 2)
5. $(p \Rightarrow q) \Rightarrow (p \Rightarrow r)$ (modus ponens' on 3,4)
6. $p \Rightarrow q$ (hypothesis)
7. $p \Rightarrow r$ (modus ponens on 5,6)

**Definition.** If $\emptyset \vdash t$, say $t$ is a *theorem*, written $\vdash t$.

**Example.** $\vdash (p \Rightarrow p)$. We want to try to get to $(p \Rightarrow (p \Rightarrow)) \Rightarrow (p \Rightarrow p)$ using axiom 2.

1. $[p \Rightarrow ((p \Rightarrow p) \Rightarrow p)] \Rightarrow [(p \Rightarrow (p \Rightarrow p)) \Rightarrow (p \Rightarrow p)]$ (axiom 2)
2. $p \Rightarrow ((p \Rightarrow p) \Rightarrow p)$ (axiom 1)
3. $(p \Rightarrow (p \Rightarrow p)) \Rightarrow (p \Rightarrow p)$ (modus ponens on 1,2)
4. $p \Rightarrow (p \Rightarrow p)$ (axiom 1)
5. $p \Rightarrow p$ (modus ponens on 3,4)

Often, showing $S \vdash p$ is made easier by:

**Proposition 1.2** (Deduction Theorem)**.** *Let $S \subseteq L$ and $p, q \in L$. Then $S \vdash (p \Rightarrow q)$ if and only if $S \cup \{p\} \vdash q$. Informally: "provability corresponds to the connective '$\Rightarrow$' in $L$".*

*Proof.* First we show ($\Rightarrow$): given a proof of $p \Rightarrow q$ from $S$, write down:

1. $p$ (hypothesis)
2. $q$ (modus ponens)

Which is a proof of $q$ from $S \cup \{p\}$.

Now we show ($\Leftarrow$): we have a proof $t_1, \dots, t_n$ of $q$ from $S \cup \{p\}$. We'll show that $S \vdash (p \Rightarrow t_i)$ for all $i$.

If $t_i$ is an axiom, write down

1. $t_i$ (axiom)
2. $t_i \Rightarrow (p \Rightarrow t_i)$ (axiom 1)
3. $p \Rightarrow t_i$ (modus ponens)

So $S \vdash (p \Rightarrow t_i)$.

If $t_i \in S$, do the same thing except step 1 will be "$t_i$ (hypothesis)" instead of "$t_i$ (axiom)".

If $t_i := p$, we have $S \vdash (p \Rightarrow p)$, since $\vdash (p \Rightarrow p)$.

If $t_i$ is obtained by modus ponens, we have $t_j$ and $t_k = (t_j \Rightarrow t_i)$ for some $j, k < n$. By induction, we can assume $S \vdash (p \Rightarrow t_j)$ and $S \vdash (p \Rightarrow (t_j \Rightarrow t_i))$. So write down

1. $[p \Rightarrow (t_j \Rightarrow t_i)] \Rightarrow [(p \Rightarrow t_j) \Rightarrow (p \Rightarrow t_i)]$ (axiom 2)
2. $(p \Rightarrow t_j) \Rightarrow (p \Rightarrow t_i)$ (modus ponens)

3. $p \Rightarrow t_i$ (modus ponens)

So $S \vdash p \Rightarrow t$. $\square$

**Example.** To show $\{p \Rightarrow q, q \Rightarrow r\} \vdash (p \Rightarrow r)$, it is sufficient to show $\{p \Rightarrow q, q \Rightarrow r, p\} \vdash r$, which is just modus ponens twice.

**Question**: how are $\vDash$ and $\vdash$ related?

**Aim**: $S \vDash t \iff S \vdash t$ (Completeness Theorem).

This is made up of:

- $S \vdash t \Rightarrow S \vDash t$ (soundness) i.e "our axioms and deduction rule are not silly";
- $S \vDash t \Rightarrow S \vdash t$ (adequacy) "our axioms are strong enough to deduce from $S$, every semantic consequence of $S$".

**Proposition 1.3** (Soundness)**.** *Let $S \subseteq L$, $t \in L$. Then $S \vdash t \Rightarrow S \vDash t$.*

*Proof.* We have a proof $t_1, \ldots, t_n$ of $t$ from $S$. So we must show that every model of $S$ is a model of $t$, i.e if $v$ is a valuation with $v(s) = 1$ for all $s \in S$, then $v(t) = 1$. But $v(p) = 1$ for each axiom $p$ (each axiom is a tautology), and for each $p \in S$ whenever $v(p) = v(p \Rightarrow q) = 1$, we have $v(q)$. So $v(t_i) = 1$ for all $i$ (induction). $\square$

One case of adequacy is: if $S \vDash \perp$, then $S \vdash \perp$. We say $S$ is *constistent* if $S \nvdash \perp$. So our statement is: $S$ has no model $\Rightarrow$ $S$ inconsistent, i.e $S$ consistent $\Rightarrow$ $S$ has a model.

In fact, this implies adequacy in general. Indeed, if $S \vDash t$ then $S \cup \{\neg t\}$ has no model. Hence (by the special case) $S \cup \{\neg t\} \vdash \perp$. So $S \vdash (\neg t \Rightarrow \perp)$, i.e $S \vdash (\neg\neg t)$. But $S \vdash (\neg\neg t) \Rightarrow t$ (axiom 3), so $S \vdash t$.

So our task is: given $S$ consistent, find a model of $S$. Could try: define

$$v(t) = \begin{cases} 1 & t \in S \\ 0 & t \notin S \end{cases}.$$

But this fails, since $S$ might not be *deductively closed*, meaning $S \vdash p \Rightarrow p \in S$. So we could first replace $S$ with its deductive closure $\{t \in L : S \vdash t\}$ (which is consistent, because $S$ is). However, this still fails: if $S$ does not 'mention' $p_3$, then $S \nvdash p_3$ and $S \nvdash \neg p_3$, so $v(p_3) = v(\neg p_3) = 0$ which is impossible.

**Theorem 1.4** (Model Existence Theorem)**.** *Let $S \subseteq L$ be consistent. Then $S$ has a model.*

Idea: extend $S$ to 'swallow up', for each $p$, one of $p$ and $\neg p$.

*Proof.* Claim: for any consistent $S \subseteq L$ and $p \in L$, $S \cup \{p\}$ or $S \cup \{\neg p\}$ is consistent.

Proof of claim: if not, then $S \cup \{p\} \vdash \perp$ and $S \cup \{\neg p\} \vdash \perp$. So $S \vdash (p \Rightarrow \perp)$ (deduction theorem), i.e $S \vdash (\neg p)$. Hence from $S \cup \{\neg p\} \vdash \perp$ we obtain $S \vdash \perp$.

Now, $L$ is countable (as each $L_n$ is countable) so we can list $L$ as $t_1, t_2, \ldots$. Let $S_0 = S$. Let $S_1 = S_0 \cup \{t_1\}$ or $S_1 \cup \{\neg t_1\}$ with $S_1$ consistent. In general, given $S_{n-1}$ let $S_n = S_{n-1} \cup \{t_n\}$ or $S_n = S_{n-1} \cup \{\neg t_n\}$ so that $S_n$ is consistent. Now set $\overline{S} = S_0 \cup S_1 \cup S_2 \cup \ldots$. Thus for all $t \in L$, either $t \in \overline{S}$ or $(\neg t) \in \overline{S}$.

Now $\overline{S}$ is consistent: if $\overline{S} \vdash \perp$ then, since proofs are finite, we'd have $S_n \vdash \perp$ for some $n$, a contradiction.

Also, $\overline{S}$ is deductively closed: if $\overline{S} \vdash p$, must have $p \in \overline{S}$, since otherwise $(\neg p) \in \overline{S}$, so $\overline{S} \vdash (p \Rightarrow \perp)$ and $\overline{S} \vdash \perp$.

Now define $v : L \to \{0, 1\}$ by

$$t \mapsto \begin{cases} 1 & t \in \overline{S} \\ 0 & \text{otherwise} \end{cases}.$$

We'll show $v$ is a valuation (then we're done as $v = 1$ on $S$).

$\underline{v(\perp)}$: have $\perp \notin \overline{S}$ (since $\overline{S}$ is consistent), so $v(\perp) = 0$.

$\underline{v(p \Rightarrow q)}$: if $v(p) = 1$, $v(q) = 0$, then have $p \in \overline{S}$, $q \notin \overline{S}$. But if $(p \Rightarrow q) \in \overline{S}$, then since $p \in \overline{S}$, $q \in \overline{S}$ (since $\overline{S}$ is deductively closed). Now if $v(q) = 1$, $q \in \overline{S}$. But $\overline{S} \vdash (q \Rightarrow (p \Rightarrow q))$ (axiom 1), so $\overline{S} \vdash (p \Rightarrow q)$ hence $(p \Rightarrow q) \in \overline{S}$ ($\overline{S}$ is deductively closed). Finally, if $v(p) = 0$, we have $p \notin \overline{S}$ and want to show $(p \Rightarrow q) \in \overline{S}$. Then $(p \Rightarrow \perp) \in \overline{S}$, so it is enough to show $(p \Rightarrow \perp) \vdash (p \Rightarrow q)$. So it's enough to show $(p, p \Rightarrow \perp) \vdash q$, so enough to show $\perp \vdash q$. But $\perp \vdash (\neg\neg q)$ (axiom 1), and $(\neg\neg q) \vdash q$ (axiom 3), so $\perp \vdash q$ as required. □

**Remarks.**

1. We used $P = (p_1, p_2, \ldots)$, in saying $L$ is countable. In fact, it also holds if $P$ is uncountable (see later in course).
2. Sometimes this theorem is called 'The Completeness Theorem'

By the remarks stated before this theorem, we have

**Corollary 1.5** (Adequacy)**.** *Let $S \subseteq L$, $t \in L$, with $S \vDash t$. Then $S \vdash t$.*

Hence we have

**Theorem 1.6** (Completeness Theorem)**.** *Let $S \subseteq L$, $t \in L$. Then $S \vdash t \iff S \vDash t$.*

**Corollary 1.7** (Compactness Theorem)**.** *Let $S \subseteq L$, $t \in L$ with $S \vDash t$. Then some finite $S' \subseteq S$ has $S' \vDash t$.*

*Proof.* This is trivial if we replace $\vDash$ by $\vdash$ (as all proofs are finite). $\square$

For $t = \perp$, the theorem says: if $S \vDash T$ then some finite $S' \subseteq S$ has $S' \vdash \perp$, i.e if every finite $S' \subseteq S$ has a model then $S$ has a model. In fact, this is equivalent to compactness in general: $S \vDash t$ says $S \cup \{\neg t\}$ has no model, and $S' \vDash t$ says $S' \cup \{\neg t\}$ has no model.

**Corollary 1.8** (Compactness Theorem equivalent form)**.** *Let $S \subseteq L$. Then if every finite subset of $S$ has a model, so does $S$.*

Another application:

**Corollary 1.9** (Decidability Theorem)**.** *Let $S \subseteq L$ be finite and $t \in L$. Then there is an algorithm to decide, in finite time, whether of not $S \vdash t$.*

**Remark.** This is a very surprising result.

*Proof.* Trivial if we replace $\vdash$ with $\vDash$: to check if $S \vDash t$ we just draw the truth table. $\square$

# 2 Well-ordering & Ordinals

**Definition.** A *total order* or *linear order* is a pair $(X, <)$ where $X$ is a set and $<$ is a relation on $X$ that is

(i) *irreflexive*: for all $x \in X$, not $x < x$;

(ii) *transitive*: for all $x, y, z \in X$, if $x < y$, $y < z$ then $x < z$;

(iii) *trichotomous*: for all $x, y \in X$, either $x = y$ or $x < y$ or $y < x$.

We sometimes write $x > y$ if $y < x$, and $x \leq y$ if $x < y$ or $x = y$.

We can instead define a total order in terms of $\leq$ as follows:

(i) *reflextive*: for all $x \in X$, $x \leq x$;

(ii) *transitive*: for all $x, y, z \in X$, if $x \leq y$, $y \leq z$ then $x \leq z$;

(iii) *antisymmetric*: for all $x, y \in X$, if $x \leq y$, $y \leq x$ then $x = y$;

(iv) *trichotomous*: for all $x, y \in X$ either $x \leq y$ or $y \leq x$.

**Examples.**

1. $\mathbb{N}, <$;
2. $\mathbb{Q}, \leq$;
3. $\mathbb{R}, \leq$;
4. $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$ under 'divides' is not a total order, e.g 2 and 3 are not related;
5. $\mathcal{P}(S), \subseteq$ is not a total order - fails trichotomy.

**Definition.** A total order $(X, <)$ is a *well-ordering* if every (non-empty) subset has a least element, i.e for all $S \subseteq X$ if $S \neq \emptyset$ then there exists $x \in S$ such that $x \leq y$ for all $y \in S$.

**Examples.**

1. $\mathbb{N}, <$;
2. $\mathbb{Z}, <$ is not a well ordering;
3. $\mathbb{Q}, <$ is not a well ordering;
4. $\mathbb{R}, <$ is not a well ordering;
5. $[0, 1] \subseteq \mathbb{R}, <$ is not a well ordering, e.g $(0, 1]$ has no least element;
6. $\{1/2, 2/3, 3/4, \ldots\} \subseteq \mathbb{R}$ is well ordered;
7. $\{1/2, 2/4, 3/4, \ldots\} \cup \{1\}$ is well ordered;

8. $\{1/2, 2/4, 3/4, \ldots\} \cup \{2\}$ is well ordered;
9. $\{1/2, 2/3, 3/4, \ldots\} \cup \{1 + 1/2, 1 + 2/3, 1 + 3/4, \ldots\}$ is well ordered.

**Remark.** $(X, <)$ is a well ordering if and only if there is no infinite strictly decreasing sequence.

We say total orders $X, Y$ are *isomorphic* if there exists a bijection $f : X \to Y$ such that $x < y$ if and only if $f(x) < f(y)$. For example, Examples 1&6, 7&8 above are isomorphic. However examples 1&7 are not isomorphic, since in 7 there exists a greatest element, but not in 1.

**Proposition 2.1** (Proof by induction)**.** *Let $X$ be well ordered and let $S \subseteq X$ be such that whenever $y \in S$ for all $y < x$, then $x \in S$. Then $S = X$. Equivalently, if $p(x)$ is a property such that $p(y)$ for all $y < x$ implies $p(x)$, then $p(x)$ for all $x \in X$.*

*Proof.* Suppose $S \neq X$ and let $x$ be least in $X \setminus S$. THen $y \in S$ for all $y < x$ but $x \notin S$, a contradiction. □

**Proposition 2.2.** *Let $X, Y$ be isomorphic well-orderings. Then there exists a unique isomorphism.*

**Note.** Note this is false for general total orders, for example $\mathbb{Z} \to \mathbb{Z}$ could have $x \mapsto x - t$ for any $t$, or $\mathbb{R} \to \mathbb{R}$ could have $x \mapsto x^3$.

*Proof.* Let $f, g : X \to Y$ be isomorphisms. We'll show $f(x) = g(x)$ for all $x$ by induction on $X$. Given $f(y) = g(y)$ for all $y < x$, we want to show $f(x) = g(x)$. We must have $f(x) = a$ where $a$ is the least element of $Y \setminus \{f(y) : y < x\}$ (non-empty since it contains $f(x)$). Indeed, if not then $f(x') = a$ for some $x' > x$, contradicting the fact $f$ is order preserving. Similarly have $g(x) = a$. □

**Definition.** A subset $I$ of a total order $X$ is an *initial segment* if $x \in I$, $y < x$ implies $y \in I$ (i.e $I$ is closed under $<$). For example $I_x = \{y \in X : y < x\}$ is an initital segment for any $x \in X$, however not every inital segment is of this form, e.g in $\mathbb{Q}$ $\{x \in \mathbb{Q} : x \leq 0 \text{ or } x^2 < 2\}$.

**Note.** In a well-ordering, every proper initial segment $I$ is of the form $I_x$, for some $x \in X$. Indeed let $x$ be the least element of $X \setminus I$ (non-empty since $I$ is proper). Then $I = I_x$, since if $y < x$ then $y \in I$ (by choice of $x$), and conversely if $y \in I$, must have $y < x$ or else $y \geq x$ implying $x \in I$ (as $I$ is an initital segment).