FIRST-ORDER LOGIC

1. First-order formulas

First-order logic is an extension of propositional logic allowing us to express statements *about elements*, instead of just pure statements. Here is an example of a first-order formula:

$$\forall x\,((0 \le x) \to \exists y\,((0 \le y) \wedge (y \cdot y = x)))$$

There are several new syntactic constructions available, compared to propositional logic:

- There are two types of expressions: formulas like $y \ge 0$, which express statements (true or false), as well as **terms** like $y \cdot y$, which denote elements rather than statements.
- Both formulas and terms may depend on **variables** like x, y. In other words, a formula $y \cdot y = x$ represents not a single truth value but rather a *relation* (binary in this case).
- **Quantifiers** like $\forall, \exists$ allow us to *bind* variables in formulas: for example, the formula $\exists y\,((0 \le y) \wedge (y \cdot y = x))$ no longer depends on y.
- There are some atomic symbols like $\le$, called **relation symbols**, to be specified in the alphabet $\mathcal{A}$, that can be used to combine terms into formulas. ($=$ can also be regarded as a binary relation symbol, although it plays a rather special role.)
- There are some other atomic symbols like $\cdot$, called **function symbols** (or **operation symbols**), also specified by $\mathcal{A}$, that can be used to combine terms into other terms. 0 above can also be regarded as a (0-ary) function symbol.

The formal definition is as follows.

1.1. **Definition.** A **first-order signature** is an alphabet $\mathcal{A}$ together with, for each $P \in \mathcal{A}$, two additional pieces of data:

- a classification of P as either a **relation symbol** or a **function symbol**;
- an **arity** $n \in \mathbb{N}$; we call P **n-ary** (or **binary** when $n = 2$, **unary** when $n = 1$, etc.).

A 0-ary (or "nullary") function symbol is also called a **constant symbol**. We write

$$\begin{aligned}
\mathcal{A}_{\mathsf{rel}} &:= \{\text{relation symbols in } \mathcal{A}\} \subseteq \mathcal{A},\\
\mathcal{A}_{\mathsf{fun}} &:= \{\text{function symbols in } \mathcal{A}\} \subseteq \mathcal{A},\\
\mathcal{A}^n_{\mathsf{rel}} &:= \{n\text{-ary relation symbols}\} \subseteq \mathcal{A}_{\mathsf{rel}},\\
\mathcal{A}^n_{\mathsf{fun}} &:= \{n\text{-ary function symbols}\} \subseteq \mathcal{A}_{\mathsf{fun}}.
\end{aligned}$$

Thus, formally, a first-order signature consists of a set $\mathcal{A}$ equipped with a partition

$$\begin{aligned}
\mathcal{A} &= \bigsqcup_{n\in\mathbb{N}} \mathcal{A}^n_{\mathsf{rel}} \sqcup \bigsqcup_{n\in\mathbb{N}} \mathcal{A}^n_{\mathsf{fun}}\\
&= \mathcal{A}^0_{\mathsf{rel}} \sqcup \mathcal{A}^1_{\mathsf{rel}} \sqcup \cdots \sqcup \mathcal{A}^0_{\mathsf{fun}} \sqcup \mathcal{A}^1_{\mathsf{fun}} \sqcup \cdots .
\end{aligned}$$

However, in practice, we usually just list out the elements of the alphabet $\mathcal{A}$, and then say in words what type of symbol each element is; for familiar symbols, like $\le, +$, we usually take them to be of the familiar type and arity.

1.2. **Example.** The **signature of graphs** is $\mathcal{A}_{\mathsf{graph}} := \{E\}$, where E is a binary relation symbol.

1.3. **Example.** The **signature of posets** is $\mathcal{A}_{\mathsf{poset}} := \{\le\}$, where $\le$ is a binary relation symbol. Note that this is identical to the signature of graphs, except for the symbol we chose to use.

1.4. **Remark.** The preceding two examples indicate an important point: *a signature can only specify what the relations/operations are*; *it cannot specify how they behave*. For example, $\mathcal{A}_{\mathsf{poset}}$ does not specify transitivity of $\leq$ in any way. In order to specify axioms that the relations/operations have to obey, we need a first-order theory (see Section 2.B below).

1.5. **Example.** The **signature of fields** is $\mathcal{A}_{\mathsf{field}} := \{+, 0, -, \cdot, 1\}$ where the symbols are, respectively, (2, 0, 1, 2, 0)-ary function symbols (so $0, 1$ are constant symbols).

(We do not include a symbol for $/$, because division is not an everywhere-defined operation; we can only require that nonzero elements in a field must have a multiplicative inverse, via an axiom in the *theory* of fields (see Example 2.31 below). Thus, this signature could just as well be called the signature of *rings* $\mathcal{A}_{\mathsf{ring}}$, illustrating again the preceding remark.)

1.6. **Example.** The **signature of ordered fields** is $\mathcal{A}_{\mathsf{ordfield}} := \mathcal{A}_{\mathsf{field}} \cup \mathcal{A}_{\mathsf{poset}} = \{+, 0, -, \cdot, 1, \leq\}$.

1.7. **Example.** The **signature of ($\mathbb{R}$-)vector spaces** is $\mathcal{A}_{\mathsf{vec}} := \{+, 0, -\} \cup \{a\cdot \mid a \in \mathbb{R}\}$, where $+, 0, -$ are as in the signature of fields above, while for each $a \in \mathbb{R}$, $a\cdot$ is a single *unary* function symbol (referring to scalar multiplication by a). So $\mathcal{A}_{\mathsf{vec}}$ is an infinite (indeed uncountable) signature.

(It would not make sense to treat $\cdot$ as a binary function symbol if we want to use this signature to describe vector spaces, since scalar multiplication does not take two vectors in a vector space V to another vector.)

1.8. **Definition.** Let $\mathcal{A}$ be a first-order signature. Fix also another alphabet $\mathcal{V}$, whose elements we call **variables**. The **$\mathcal{A}$-terms with variables from $\mathcal{V}$** are constructed inductively as follows:

- Every $x \in \mathcal{V}$ is an $\mathcal{A}$-term.
- If $f \in \mathcal{A}^n_{\mathsf{fun}}$ is an n-ary function symbol, and $t_1, \ldots, t_n$ are terms, then so is $f(t_1, \ldots, t_n)$.

The **(first-order) $\mathcal{A}$-formulas with variables from $\mathcal{V}$** are constructed inductively as follows:

- If $R \in \mathcal{A}^n_{\mathsf{rel}}$ is an n-ary relation symbol, or the symbol $=$ when $n = 2$, and $t_1, \ldots, t_n$ are $\mathcal{A}$-terms, then $R(t_1, \ldots, t_n)$ is an $\mathcal{A}$-formula, called an **atomic formula**.[1]
- If ϕ, ψ are $\mathcal{A}$-formulas, then $\phi \wedge \psi, \phi \vee \psi, \neg\phi$ are $\mathcal{A}$-formulas.
- $\top, \bot$ are $\mathcal{A}$-formulas.
- If ϕ is an $\mathcal{A}$-formula, and $x \in \mathcal{V}$ is a variable, then $\exists x\, \phi$ is an $\mathcal{A}$-formula.

We continue to use the abbreviations $\rightarrow, \leftrightarrow$ as in propositional logic, as well as

$$\forall x\, \phi := \neg\exists x\, \neg\phi.$$

(The reason for regarding $\forall$ as an abbreviation, rather than $\exists$, is similar to why we chose to regard $\rightarrow$ as an abbreviation in propositional logic, but not $\phi \vee \psi := \neg(\neg\phi \wedge \neg\psi)$, say: we will use them to illustrate different aspects of our proof system for first-order logic. Indeed, there is a sense in which $\forall$ is analogous to $\rightarrow$ and $\exists$ to $\vee$; see Remark 3.18 and Example 3.26 below.)

1.9. **Example.** Let $x, y \in \mathcal{V}$ be variables. The following is an $\mathcal{A}_{\mathsf{ordfield}}$-term:

$$+(1, \cdot(x, y))$$

When we are dealing with signatures consisting of familiar symbols like $+, \cdot$, we will write terms and formulas in the familiar way; e.g., the above term would usually be written

$$1 + x \cdot y.$$

Likewise, the $\mathcal{A}_{\mathsf{ordfield}}$-formula given at the beginning of this section is a more familiar way of writing

$$\forall x\, (\leq(0, x) \rightarrow \exists y\, (\leq(0, y) \wedge =(\cdot(y, y), x))).$$

[1] There is an abuse of notation going on here: for two terms s, t, "$s = t$" may denote *either* the "meta" assertion that these two terms are the same (as expressions), *or* the atomic formula $s = t$! Some logic books therefore use a different symbol (like $\equiv$) for the equality formula. We will instead depend on context to disambiguate.

1.10. **Example.** The following are *not* $\mathcal{A}_{\mathsf{ordfield}}$-formulas:

$\leq(x, y, z)$	($\leq$ is binary, not ternary)
$\forall x\,(x + y \cdot y)$	($\forall x$ must be followed by a formula, not a term)
$\forall x\,(\bot \leq x \cdot x)$	(the LHS of $\leq$ must be a term, not a formula)
$\forall x\,(\sqrt{x} \cdot \sqrt{x} = x)$	(no $\sqrt{\ }$ symbol in $\mathcal{A}_{\mathsf{ordfield}}$)
$\forall x\,(2 + x = x + 2)$	(no 2 symbol in $\mathcal{A}_{\mathsf{ordfield}}$)

However, we might treat the last formula as an abbreviation for

$$\forall x\,((1 + 1) + x = x + (1 + 1)).$$

On the other hand, the following *are* $\mathcal{A}_{\mathsf{ordfield}}$-formulas:

$0 = 1$	(will be interpreted as false)
$\exists x\,\top$	(will be interpreted as "the model is nonempty")
$0 \leq x \to \forall x\,\exists x\,(x \leq 0)$	(nothing in definition of formula prevents variable clashes)

1.A. **Free and bound variables.** The last example above shows that in order to interpret formulas correctly, it is important to pay attention to which variables occur underneath quantifiers.

An occurrence of a variable underneath a quantifier in a formula is called **bound**; a **free variable** in a formula is a variable which occurs non-bound (at least once). Since terms do not contain quantifiers, all variables in terms are considered free. Formally:

1.11. **Definition.** The **set of free variables** $\mathrm{FV}(t), \mathrm{FV}(\phi)$ of a term t or formula ϕ is defined inductively as follows:

$$\begin{aligned}
\mathrm{FV}(x) &:= \{x\},\\
\mathrm{FV}(f(t_1, \dots, t_n)) &:= \mathrm{FV}(t_1) \cup \dots \cup \mathrm{FV}(t_n) \quad \text{for } f \in \mathcal{A}^n_{\mathsf{fun}} \text{ and terms } t_1, \dots, t_n,\\
\mathrm{FV}(R(t_1, \dots, t_n)) &:= \mathrm{FV}(t_1) \cup \dots \cup \mathrm{FV}(t_n) \quad \text{for } R \in \mathcal{A}^n_{\mathsf{rel}} \text{ (or } R = =) \text{ and terms } t_1, \dots, t_n,\\
\mathrm{FV}(\phi \wedge \psi) &:= \mathrm{FV}(\phi \vee \psi) := \mathrm{FV}(\phi) \cup \mathrm{FV}(\psi),\\
\mathrm{FV}(\neg\phi) &:= \mathrm{FV}(\phi),\\
\mathrm{FV}(\top) &:= \mathrm{FV}(\bot) := \varnothing,\\
\mathrm{FV}(\exists x\,\phi) &:= \mathrm{FV}(\phi) \setminus \{x\}.
\end{aligned}$$

(Compare with the definition of $\mathrm{AT}(\phi)$ in Example 1.14 from propositional logic.) It follows that

$$\begin{aligned}
\mathrm{FV}(\phi \to \psi) &= \mathrm{FV}(\neg\phi \vee \psi) = \mathrm{FV}(\phi) \cup \mathrm{FV}(\psi),\\
\mathrm{FV}(\phi \leftrightarrow \psi) &= \mathrm{FV}(\phi) \cup \mathrm{FV}(\psi),\\
\mathrm{FV}(\forall x\,\phi) &= \mathrm{FV}(\neg\exists x\,\neg\phi) = \mathrm{FV}(\phi) \setminus \{x\}.
\end{aligned}$$

A formula is called a **sentence** if it has no free variables.

1.12. **Example.** To compute the free variables of the $\mathcal{A}_{\mathsf{ordfield}}$-formula from page 1:

$$\underbrace{\forall x\,\underbrace{(\overbrace{(0 \leq x)}^{\mathrm{FV}=\{x\}} \to \exists y\,\underbrace{(\overbrace{(0 \leq y)}^{\mathrm{FV}=\{y\}} \wedge \overbrace{(y \cdot y = x)}^{\mathrm{FV}=\{x,y\}})}_{\mathrm{FV}=\{y\}\cup\{x,y\}=\{x,y\}})}_{\mathrm{FV}=\{x,y\}\setminus\{y\}=\{x\}}}_{\mathrm{FV}=\{x\}\cup\{x\}=\{x\}}}_{\mathrm{FV}=\{x\}\setminus\{x\}=\varnothing}$$

1.13. **Exercise.** Compute the free variables of the following formulas:

(a) $(0 \leq x) \to \forall x\, \exists x\, (x \leq 0)$
(b) $(\forall x\, (x \leq x \cdot y)) \vee (\exists y\, (x \cdot y \leq y))$
(c) $\forall x\, ((\forall y\, (x \leq y \cdot z)) \to \exists x\, (x + y = z))$

The set of *free* variables in a formula is much more important than what *all* the variables are, since bound variables may be changed without affecting the meaning of the formula: for example,

$$\exists y\, (x + y = 0) \text{ vs. } \exists z\, (x + z = 0)$$

should always have the same meaning. (Of course, as always, these two formulas are not quite *equal*, but only "equal mod bound variables", or *α-equivalent*; see the Appendix on variable substitution.) For this reason, from now on, we will never mention the set $\mathcal{V}$ from which all variables are drawn; we will only ever care what the free variables of a formula are. For any set of variables X, we write

$$\mathcal{L}^X_{\mathsf{term}}(\mathcal{A}) := \{\mathcal{A}\text{-terms } t \mid \mathrm{FV}(t) \subseteq X\},$$
$$\mathcal{L}^X_{\mathsf{form}}(\mathcal{A}) := \{\mathcal{A}\text{-formulas } \phi \mid \mathrm{FV}(\phi) \subseteq X\},$$

and call these the set of t, ϕ respectively with **free variables from** X.[2]

1.14. **Remark.** It is important to note that when we say ϕ has free variables from X, we do not actually require each $x \in X$ to occur in ϕ; we only care that no other variables can occur free in ϕ. This is usually more important than knowing which free variables actually do occur: for example, as long as ϕ has free variables from $\{x, y\}$, then $\forall x\, \exists y\, \phi$ will be a sentence.

2. First-order semantics

2.1. **Definition.** Let $\mathcal{A}$ be a first-order signature. An **$\mathcal{A}$-structure** $\mathcal{M}$ consists of:

- an **underlying set** (also called **domain** or **universe**), denoted M or $|\mathcal{M}|$;
- for each n-ary relation symbol $R \in \mathcal{A}^n_{\mathsf{rel}}$, an n-ary relation $R^{\mathcal{M}} = \mathcal{M}(R)$ on M;
- for each n-ary function symbol $f \in \mathcal{A}^n_{\mathsf{fun}}$, an n-ary function $f^{\mathcal{M}} = \mathcal{M}(f) : M^n \to M$.

We call $R^{\mathcal{M}}, f^{\mathcal{M}}$ the **interpretation of** R, f **in** $\mathcal{M}$.

Here M^n denotes the n-fold Cartesian product (set of n-tuples)

$$M^n := M \times \cdots \times M = \{(a_1, \ldots, a_n) \mid a_1, \ldots, a_n \in M\}.$$

When $n = 0$, $M^0 = \{()\}$ is a one-element set consisting of the empty tuple $()$. An 0-ary function $f : M^0 \to M$ thus consists of simply an element $f(()) \in M$; we usually identify f with $f(())$. Thus, each constant symbol $c \in \mathcal{A}^0_{\mathsf{fun}}$ is interpreted as simply an element $c^{\mathcal{M}} \in M$.

An **n-ary relation** R on M may be represented in multiple ways:

- R may be thought of as a subset $R \subseteq M^n$, namely the set of all n-tuples at which R holds. For example, the equality relation on M is represented as

$$\begin{aligned}(=_M) &= \{(a, b) \in M^2 \mid a = b\} \subseteq M^2 \\ &= \{(a, a) \mid a \in M\}.\end{aligned}$$

  To say that R **holds at a tuple** $\vec{a} = (a_1, \ldots, a_n)$ then means that $\vec{a} \in R$.

[2] A terrible but very common abuse of notation that one sees in first-order logic is to write $\phi(x, y, \ldots)$ for a formula with free variables among $x, y, \ldots$, and then write $\phi(a, b, \ldots)$ when "plugging in" a, b for x, y (either syntactically, via substitution as in the Appendix, or semantically, as in Definition 2.13). This is terrible because the mathematical object ϕ does not "know" that x is its first variable, y is its second, etc.; formally, it does not make sense to distinguish between $\phi(x, y)$ and $\phi(y, x)$. It is best to avoid this notation in any serious discussion of the syntax of formulas.

- R may also be thought of as a function $R : M^n \to \{0,1\}$, which specifies whether or not R holds at each n-tuple. For example, the equality relation on M is represented as

$$(=_M) : M^2 \longrightarrow \{0,1\}$$
$$(a,b) \longmapsto \begin{cases} 1 & \text{if } a = b, \\ 0 & \text{if } a \neq b. \end{cases}$$

  To say that R **holds at** $\vec{a}$ then means that $R(\vec{a}) = 1$.

Given a subset $R \subseteq M^n$, the corresponding function to $\{0,1\}$ is its **indicator function** (or **characteristic function**)

$$\mathbb{1}_R : M^n \longrightarrow \{0,1\}$$
$$\vec{a} \longmapsto \begin{cases} 1 & \text{if } \vec{a} \in R, \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

The correspondence $R \mapsto \mathbb{1}_R$ is a bijection between the set $\mathcal{P}(M^n)$ of all subsets of M^n, and the set $\{0,1\}^{M^n}$ of all functions $M^n \to \{0,1\}$ (see the Appendix); thus we may think of either of these as representing the relation R. *By an abuse of notation, we will henceforth identify these two representations.* Roughly speaking, they correspond to the "satisfaction predicate $m \models P$" versus "truth value $m(P)$" views of propositional semantics, which are both convenient in different contexts.

2.3. **Example.** We have a $\mathcal{A}_{\mathsf{ordfield}}$-structure $\mathcal{R}$ with underlying set $\mathbb{R}$ and each symbol in $\mathcal{A}_{\mathsf{ordfield}}$ interpreted as the usual operation or relation of that name, e.g., $+^{\mathcal{R}} : \mathbb{R}^2 \to \mathbb{R}$ is the binary addition function. Similarly, we have a $\mathcal{A}_{\mathsf{ordfield}}$-structure $\mathcal{Q}$ consisting of $\mathbb{Q}$ and the usual interpretations.

2.4. **Example.** We have a $\mathcal{A}_{\mathsf{ordfield}}$-structure $\mathcal{M}$ with underlying set $\mathbb{R}$ and

$$+^{\mathcal{M}} := \text{usual } +,$$
$$0^{\mathcal{M}} := \text{usual } 0,$$
$$-^{\mathcal{M}} := \text{usual sin},$$
$$\cdot^{\mathcal{M}} := \text{usual } +,$$
$$1^{\mathcal{M}} := \text{usual } \pi,$$
$$\leq^{\mathcal{M}} := \text{usual } =.$$

(Nothing in the definition of $\mathcal{A}_{\mathsf{ordfield}}$-structure says that the field axioms like commutativity, etc., have to hold; this will be enforced by the first-order *theory* of fields, see Example 2.31.)

2.5. **Example.** Similarly, a $\mathcal{A}_{\mathsf{poset}}$-structure $\mathcal{M}$ is a set M equipped with an *arbitrary* binary relation $\leq^{\mathcal{M}} \subseteq M^2$ (not yet required to be a partial order).

2.6. **Example.** For $\mathcal{A} = \varnothing$, an $\mathcal{A}$-structure is just a set.

2.A. **Interpretation of terms and formulas.** Let $\mathcal{M}$ be an $\mathcal{A}$-structure. In order to interpret a term or formula in $\mathcal{M}$, we need to know what values are assigned to its free variables; in other words, the interpretation will be a *function* defined on the set of all variable assignments (to either M or $\{0,1\}$, depending on whether we have a term or a formula).

By a **variable assignment** in M, we just mean a function $\alpha : X \to M$ from a set of variables X. The set of all X-indexed variable assignments is thus M^X, the set of all functions from X to M. Note that we can also think of $\alpha : X \to M$ as an "X-ary tuple" of elements of M, namely $(\alpha(x))_{x \in X}$; this allows us to think of the interpretation of terms and formulas as generalizing the interpretation of function and relation symbols in $\mathcal{A}$ (which yield functions $M^n \to M$ or $M^n \to \{0,1\}$).

2.7. **Definition.** We begin with the interpretation of terms. For each $\mathcal{A}$-term $t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$ with free variables from some set X, we will define by induction on t a function

$$t^{\mathcal{M}} = t^{\mathcal{M}}_X = \mathcal{M}_X(t) : M^X \longrightarrow M,$$

called the **interpretation of t in $\mathcal{M}$**, which maps each variable assignment $\alpha \in M^X$ to an element $t^{\mathcal{M}}_X(\alpha) \in M$, called the **interpretation of t in $\mathcal{M}$ under the variable assignment α**:

- For a single variable $x \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$ with free variables from X, this means $x \in X$; we define
$$x^{\mathcal{M}}(\alpha) := \alpha(x).$$
- For a term $f(t_1, \ldots, t_n) \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$ where $f \in \mathcal{A}^n_{\mathsf{fun}}$ and $t_1, \ldots, t_n \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$, we define
$$f(t_1, \ldots, t_n)^{\mathcal{M}}(\alpha) := f^{\mathcal{M}}(t_1^{\mathcal{M}}(\alpha), \ldots, t_n^{\mathcal{M}}(\alpha))$$
(recall that $f^{\mathcal{M}} : M^n \to M$ is provided as part of the structure $\mathcal{M}$).

2.8. **Example.** In the $\mathcal{A}_{\mathsf{ordfield}}$-structure $\mathcal{M} = \mathbb{R}$ with weird operations from Example 2.4,

$$\begin{aligned}
&((-1)\cdot(x+y))^{\mathcal{M}}(x \mapsto 3,\, y \mapsto 5) \\
&\quad = (-1)^{\mathcal{M}}(x \mapsto 3,\, y \mapsto 5) \cdot^{\mathcal{M}} (x+y)^{\mathcal{M}}(x \mapsto 3,\, y \mapsto 5) \\
&\quad = (-^{\mathcal{M}}1^{\mathcal{M}}(x \mapsto 3,\, y \mapsto 5)) + (x^{\mathcal{M}}(x \mapsto 3,\, y \mapsto 5) +^{\mathcal{M}} y^{\mathcal{M}}(x \mapsto 3,\, y \mapsto 5)) \\
&\quad = \sin(\pi) + (3+5) = 8.
\end{aligned}$$

2.9. **Exercise.** Verify that this is the same as $((-1)\cdot x + (-1)\cdot y)^{\mathcal{M}}(x \mapsto 3,\, y \mapsto 5)$.
Is it also the same as $((-x)+(-y))^{\mathcal{M}}(x \mapsto 3,\, y \mapsto 5)$?

2.10. **Remark.** You will probably have noticed all of the redundant variables we had to keep writing in the above example. Strictly speaking, this is necessary, since the inductive case of Definition 2.7 for e.g., $(x+y)^{\mathcal{M}}$ refers to $x^{\mathcal{M}}$ and $y^{\mathcal{M}}$ with respect to the *same* set of free variables $\{x, y\}$. Indeed, strictly speaking the notation $x^{\mathcal{M}}$ is ambiguous in the absence of a variable assignment (as in $x^{\mathcal{M}}(x \mapsto 3,\, y \mapsto 5)$): it could mean $x^{\mathcal{M}}_{\{x,y\}} : M^{\{x,y\}} \to M$, or $x^{\mathcal{M}}_{\{x\}} : M^{\{x\}} \to M$, *which are entirely different functions with different domains*. Hence, officially we should include the set of variables in the subscript, for clarity.

Fortunately, the following shows that when the term or formula does not mention all of the available free variables, then the interpretation under a *particular* variable assignment α stays the same when we drop the extraneous variables. Thus for example, above it would've been safe to write $(-1)^{\mathcal{M}}()$ instead of $(-1)^{\mathcal{M}}(x \mapsto 3,\, y \mapsto 5)$, and $x^{\mathcal{M}}(x \mapsto 3)$ instead of $x^{\mathcal{M}}(x \mapsto 3,\, y \mapsto 5)$, etc.

2.11. **Exercise.** Let $\alpha : Y \to M$ be a variable assignment and $X \subseteq Y$.

(a) If a term t only has free variables from X, then $t^{\mathcal{M}}_X(\alpha|X) = t^{\mathcal{M}}_Y(\alpha)$.
(b) If a formula ϕ only has free variables from X, then $\phi^{\mathcal{M}}_X(\alpha|X) = \phi^{\mathcal{M}}_Y(\alpha)$ [as defined below].

2.12. **Definition.** When interpreting formulas, in the $\exists$ case, because the subformula has one more free variable, we will need to extend our given variable assignment to include that extra variable. We therefore introduce the following general notation: for a function $\alpha : X \to M$, a variable x (which may or may not be in X), and an element $a \in M$,

$$\begin{aligned}
\alpha\langle x \mapsto a\rangle : X \cup \{x\} &\longrightarrow M \\
y &\longmapsto \begin{cases} a & \text{if } y = x, \\ \alpha(y) & \text{if } y \in X \setminus \{x\}. \end{cases}
\end{aligned}$$

In other words, we add the assignment $x \mapsto a$ to α, replacing the previous value of $\alpha(x)$ if any.

2.13. **Definition.** We now give the interpretation of formulas. For an $\mathcal{A}$-formula $\phi \in \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$ with free variables from X, its **interpretation** $\phi^{\mathcal{M}}_X = \mathcal{M}_X(\phi)$ **in** $\mathcal{M}$ will be an "X-ary relation" on M, hence as noted above (2.2) may be represented as either a set of "X-ary tuples"

$$\phi^{\mathcal{M}} = \phi^{\mathcal{M}}_X = \mathcal{M}_X(\phi) \subseteq M^X$$

or its indicator function

$$\phi^{\mathcal{M}} = \phi^{\mathcal{M}}_X = \mathcal{M}_X(\phi) : M^X \longrightarrow \{0,1\}.$$

If $\phi^{\mathcal{M}}(\alpha) = 1$, i.e., $\alpha \in \phi^{\mathcal{M}}_X$, then we say that $\mathcal{M}$ **satisfies** ϕ **under** α, also denoted

$$\mathcal{M} \models_\alpha \phi \;:\iff\; \phi^{\mathcal{M}}(\alpha) = 1 \;\iff\; \alpha \in \phi^{\mathcal{M}}_X.$$

We will give the inductive definition using all three of these equivalent notations at once.

- For $\phi = R(t_1, \dots, t_n) \in \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$ where $R \in \mathcal{A}^n_{\mathsf{rel}}$ and $t_1, \dots, t_n \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$, we define

  $$R(t_1, \dots, t_n)^{\mathcal{M}}(\alpha) := R^{\mathcal{M}}(t_1^{\mathcal{M}}(\alpha), \dots, t_n^{\mathcal{M}}(\alpha)),$$

  similarly to the inductive case for terms in Definition 2.7. Equivalently,

  $$\mathcal{M} \models_\alpha R(t_1, \dots, t_n) \;:\iff\; (t_1^{\mathcal{M}}(\alpha), \dots, t_n^{\mathcal{M}}(\alpha)) \in R^{\mathcal{M}},$$
  $$R(t_1, \dots, t_n)^{\mathcal{M}} := (t_1^{\mathcal{M}}, \dots, t_n^{\mathcal{M}})^{-1}(R^{\mathcal{M}})$$

  where the RHS on the last line denotes the preimage of $R^{\mathcal{M}} \subseteq M^n$ under the function $(t_1^{\mathcal{M}}, \dots, t_n^{\mathcal{M}}) : M^X \to M^n$ whose coordinates are the $t_i^{\mathcal{M}} : M^X \to M$.

  When R is the equality symbol $=$, we always take $=^{\mathcal{M}}$ to be the equality relation, i.e., the set or function $=_M$ defined in the discussion before (2.2), or equivalently,

  $$\mathcal{M} \models_\alpha s = t \;:\iff\; s^{\mathcal{M}}(\alpha) = t^{\mathcal{M}}(\alpha).$$

- The connective cases are the same as in propositional logic: for $\phi, \psi \in \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$,

  $$(\phi \wedge \psi)^{\mathcal{M}}(\alpha) := \min(\phi^{\mathcal{M}}(\alpha), \psi^{\mathcal{M}}(\alpha)),$$
  $$(\phi \vee \psi)^{\mathcal{M}}(\alpha) := \max(\phi^{\mathcal{M}}(\alpha), \psi^{\mathcal{M}}(\alpha)),$$
  $$(\neg\phi)^{\mathcal{M}}(\alpha) := 1 - \phi^{\mathcal{M}}(\alpha),$$
  $$\top^{\mathcal{M}}(\alpha) := 1,$$
  $$\bot^{\mathcal{M}}(\alpha) := 0.$$

  Equivalently,

  $$\mathcal{M} \models_\alpha \phi \wedge \psi \;:\iff\; \mathcal{M} \models_\alpha \phi \text{ and } \mathcal{M} \models_\alpha \psi, \qquad (\phi \wedge \psi)^{\mathcal{M}} := \phi^{\mathcal{M}} \cap \psi^{\mathcal{M}},$$
  $$\mathcal{M} \models_\alpha \phi \vee \psi \;:\iff\; \mathcal{M} \models_\alpha \phi \text{ or } \mathcal{M} \models_\alpha \psi, \qquad (\phi \vee \psi)^{\mathcal{M}} := \phi^{\mathcal{M}} \cup \psi^{\mathcal{M}},$$
  $$\mathcal{M} \models_\alpha \neg\phi \;:\iff\; \mathcal{M} \not\models_\alpha \phi, \qquad (\neg\phi)^{\mathcal{M}} := M^X \setminus \phi^{\mathcal{M}},$$
  $$\mathcal{M} \models_\alpha \top \quad \text{always}, \qquad \top^{\mathcal{M}} := M^X,$$
  $$\mathcal{M} \models_\alpha \bot \quad \text{never}, \qquad \bot^{\mathcal{M}} := \varnothing.$$

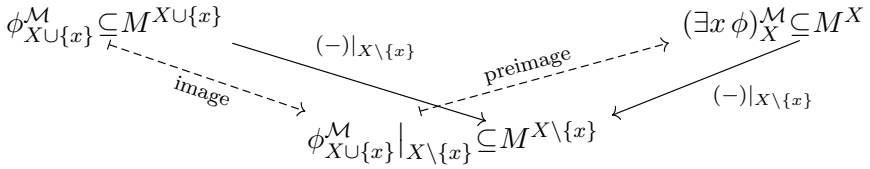
- Finally, suppose $\exists x\, \phi \in \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$; then from Definition 1.11 of free variables, we have $X \supseteq \mathrm{FV}(\exists x\, \phi) = \mathrm{FV}(\phi) \setminus \{x\}$, whence $X \cup \{x\} \supseteq \mathrm{FV}(\phi)$, so that (by the IH) we may assume given the interpretation of ϕ under any $(X \cup \{x\})$-variable assignment. We then define

  $$(\exists x\, \phi)^{\mathcal{M}}_X(\alpha) := \max_{a \in M} \phi^{\mathcal{M}}_{X \cup \{x\}}(\alpha\langle x \mapsto a\rangle)$$

  (where by convention, the max is 0 if $M = \varnothing$). In other words, we interpret $\exists x\, \phi$ as true iff there is some a we can assign to x (ignoring any previous assignment in α) to make ϕ true:

  $$\mathcal{M} \models_\alpha \exists x\, \phi \;:\iff\; \text{there exists } a \in M \text{ s.t. } \mathcal{M} \models_{\alpha\langle x \mapsto a\rangle} \phi.$$

Only read this paragraph if you're feeling brave. The definition of $(\exists x\,\phi)^{\mathcal{M}}_X \subseteq M^X$ as a set is a bit more involved to describe. Start with $\phi^{\mathcal{M}}_{X\cup\{x\}} \subseteq M^{X\cup\{x\}}$, and consider the restriction map $(-)|_{X\setminus\{x\}} : M^{X\cup\{x\}} \to M^{X\setminus\{x\}}$ which forgets about the value of a variable assignment at x; then the image $\phi^{\mathcal{M}}_{X\cup\{x\}}\big|_{X\setminus\{x\}} \subseteq M^{X\setminus\{x\}}$ is the set of assignments which can be extended to x satisfying ϕ. But since we also need to ignore any previous assignment to x, consider also the restriction $(-)|_{X\setminus\{x\}} : M^X \to M^{X\setminus\{x\}}$ (which is either the identity function if $x \notin X$, or otherwise is the same as the former restriction); then the *preimage* of $\phi^{\mathcal{M}}_{X\cup\{x\}}\big|_{X\setminus\{x\}}$ is the set of assignments for which we can first discard any previous assignment to x, and *then* extend with a new assignment to x satisfying ϕ, which exactly yields $(\exists x\,\phi)^{\mathcal{M}}_X$. The following diagram depicts this two-step construction of $(\exists x\,\phi)^{\mathcal{M}}_X$ from $\phi^{\mathcal{M}}_{X\cup\{x\}}$:



- Let us also record the interpretation of $\forall x\,\phi := \neg\exists x\,\neg\phi$, derived from that of $\exists, \neg$:

$$(\forall x\,\phi)^{\mathcal{M}}_X(\alpha) := \min_{a\in M} \phi^{\mathcal{M}}_{X\cup\{x\}}(\alpha\langle x \mapsto a\rangle),$$
$$\mathcal{M} \models_\alpha \forall x\,\phi \iff \text{for all } a \in M,\ \mathcal{M} \models_{\alpha\langle x\mapsto a\rangle} \phi.$$

(The definition of $(\forall x\,\phi)^{\mathcal{M}}_X \subseteq M^X$ as a set is similar to that of $(\exists x\,\phi)^{\mathcal{M}}_X$, involving a "coimage" rather than an image, and is left to you as an Exercise.)

2.14. **Example.** In the $\mathcal{A}_{\mathsf{ordfield}}$-structure $\mathcal{R} := \mathbb{R}$ with the usual interpretations, consider the sentence $\forall x\,\exists y\,(\neg(y = x) \wedge (x \le y))$ under the empty variable assignment:

$$\begin{aligned}
&\mathcal{R} \models_\varnothing \forall x\,\exists y\,(\neg(y = x) \wedge (x \le y))\\
&\iff \forall a \in \mathbb{R},\ \mathcal{R} \models_{x\mapsto a} \exists y\,(\neg(y = x) \wedge (x \le y))\\
&\iff \forall a \in \mathbb{R},\ \exists b \in \mathbb{R} \text{ s.t. } \mathcal{R} \models_{x\mapsto a, y\mapsto b} \neg(y = x) \wedge (x \le y)\\
&\iff \forall a \in \mathbb{R},\ \exists b \in \mathbb{R} \text{ s.t. } (\mathcal{R} \not\models_{x\mapsto a, y\mapsto b} y = x \text{ and } \mathcal{R} \models_{x\mapsto a, y\mapsto b} x \le y)\\
&\iff \forall a \in \mathbb{R},\ \exists b \in \mathbb{R} \text{ s.t. } (b \ne a \text{ and } a \le b)\\
&\iff \forall a \in \mathbb{R},\ \exists b \in \mathbb{R} \text{ s.t. } a < b
\end{aligned}$$

which is clearly true, since given a, we can take $b := a + 1$.

We can also regard $\forall x\,\exists y\,(\neg(y = x) \wedge x \le y)$ as having free variables from $\{x, y\}$, hence interpret it under some assignment of those variables, e.g.,

$$\begin{aligned}
&\mathcal{R} \models_{x\mapsto 3, y\mapsto 2} \forall x\,\exists y\,(\neg(y = x) \wedge (x \le y))\\
&\iff \forall a \in \mathbb{R},\ \mathcal{R} \models_{x\mapsto a, y\mapsto 2} \exists y\,(\neg(y = x) \wedge (x \le y))\\
&\iff \forall a \in \mathbb{R},\ \exists b \in \mathbb{R} \text{ s.t. } \mathcal{R} \models_{x\mapsto a, y\mapsto b} \neg(y = x) \wedge (x \le y)\\
&\iff \forall a \in \mathbb{R},\ \exists b \in \mathbb{R} \text{ s.t. } (\mathcal{R} \not\models_{x\mapsto a, y\mapsto b} y = x \text{ and } \mathcal{R} \models_{x\mapsto a, y\mapsto b} x \le y)\\
&\iff \forall a \in \mathbb{R},\ \exists b \in \mathbb{R} \text{ s.t. } (b \ne a \text{ and } a \le b)\\
&\iff \forall a \in \mathbb{R},\ \exists b \in \mathbb{R} \text{ s.t. } a < b
\end{aligned}$$

which is true as before. Note that the original values of x, y are "overridden": e.g., in the second line, we used

$$(x \mapsto 3, y \mapsto 2)\langle x \mapsto a\rangle = (x \mapsto a, y \mapsto 2).$$

In particular, the formula $x \le y$ is never evaluated with the original values $x = 3$, $y = 2$.

2.15. **Exercise.** Determine whether or not $\mathcal{R} \models_{x\mapsto 1, y\mapsto 2} \forall x\,((\exists x\,(y \cdot y = x)) \to (\exists y\,(y \cdot y = x)))$.

2.16. **Exercise.** Let ϕ be any formula with free variables from X, and let $x \in X$. Verify that for any $\mathcal{M}$ and $\alpha : X \to M$,

$$\mathcal{M} \models_\alpha (\forall x\,\phi) \to \phi.$$

(This reflects the common situation where you find yourself knowing that "for all x, ...", and you use this to conclude that ... holds for an already fixed x. We will see the inference rule that formalizes this way of reasoning in Example 3.26 below.) What about

$$\mathcal{M} \models_\alpha (\exists x\,\phi) \to \phi,$$
$$\mathcal{M} \models_\alpha \phi \to (\forall x\,\phi),$$
$$\mathcal{M} \models_\alpha \phi \to (\exists x\,\phi)?$$

2.17. **Definition.** We call $\phi \in \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$ a **semantic tautology (over X)**, written

$$\models_X \phi,$$

if for every $\mathcal{A}$-structure $\mathcal{M}$ and variable assignment $\alpha : X \to M$, we have $\mathcal{M} \models_\alpha \phi$. If $X = \varnothing$, we drop the subscript. (When we don't say "over X", we typically mean over $\mathrm{FV}(\phi)$.)

We say $\phi \in \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$ **semantically implies** $\psi \in \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$ (over X) if

$$\phi \models_X \psi \;:\iff\; \models_X \phi \to \psi,$$

and that they are **semantically equivalent (over X)** if

$$\phi \dashv\models_X \psi \;:\iff\; \phi \models_X \psi \text{ and } \psi \models_X \phi.$$

Finally, we say $\phi \in \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$ is **satisfiable (over X)** if for *some* $\mathcal{A}$-structure $\mathcal{M}$ and variable assignment $\alpha : X \to M$, we have $\mathcal{M} \models_\alpha \phi$. Thus

$$\phi \text{ satisfiable over } X \iff \not\models_X \neg\phi.$$

2.18. **Example.** $\forall x\,\forall y\,\forall z\,((x = y) \wedge (y = z) \to (x = z))$ is a semantic tautology, since for any $\mathcal{M}$,

$$\mathcal{M} \models \forall x\,\forall y\,\forall z\,((x = y) \wedge (y = z) \to (x = z)) \iff \forall a, b, c \in M\,(a = b \text{ and } b = c \implies a = c)$$

which is true by transitivity of equality. (So $(x = y) \wedge (y = z)$ semantically implies $x = z$.)

2.19. **Example.** $(x + y) + z = x + (y + z)$ is *not* a semantic tautology, since its interpretation in $\mathcal{M} := \mathbb{R}$ with $+^{\mathcal{M}} :=$ subtraction is false, under the assignment $x \mapsto 0$, $y \mapsto 1$, and $z \mapsto 1$, say.

2.20. **Remark.** The "over X" part of the definition of "semantic tautology" is technically necessary: if ϕ has free variables from $X \subseteq Y$, then whether ϕ is a semantic tautology could depend on whether we regard it has having free variables from X or from Y; that is, we might not have $\models_X \phi \iff \models_Y \phi$.

Indeed, consider the sentence $\phi := \exists x\,\top$ (mentioned in Example 1.10). For any $\mathcal{M}$, we have

$$\begin{aligned}\mathcal{M} \models \phi &\iff \exists a \in M \text{ s.t. } \mathcal{M} \models_{x\mapsto a} \top \\ &\iff \exists a \in M,\end{aligned}$$

i.e., ϕ asserts that the underlying set is nonempty. So ϕ is not a tautology *as a sentence*, i.e., $\not\models_\varnothing \phi$, since ϕ is false in an empty structure. However, if we instead regard ϕ as having free variables from some *nonempty* X, then ϕ *is* a tautology, i.e., $\models_X \phi$, since for any $\mathcal{M}$ and variable assignment $\alpha \in M^X$, since X is nonempty, M must be nonempty, whence $\mathcal{M} \models \phi$.

However, the following shows that this issue really does only come up in edge cases:

2.21. **Exercise.**

(a) Show that if $X \subseteq Y$, then $\models_X \phi \implies \models_Y \phi$.

(b) Show that the converse holds assuming either $X \neq \varnothing$, or every empty structure satisfies ϕ.

2.B. **Theories.** Let $\mathcal{A}$ be a first-order signature, X be a set of variables. An **$\mathcal{A}$-theory with free variables from** X is an arbitrary set of formulas $\mathcal{T} \subseteq \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$, called **axioms** of $\mathcal{T}$. When we just say **$\mathcal{A}$-theory**, we typically mean without free variables, i.e., a set of sentences. These latter ones are the most important kinds of theories that show up in concrete examples; theories with free variables serve a more theoretical purpose (see Remark 2.40 and Proposition 3.37).

For an $\mathcal{A}$-theory (without free variables) $\mathcal{T}$, a **model of** $\mathcal{T}$ is an $\mathcal{A}$-structure $\mathcal{M}$ which satisfies every axiom of $\mathcal{T}$, denoted

$$\mathcal{M} \models \mathcal{T} \;:\Longleftrightarrow\; \forall \phi \in \mathcal{T}\, (\mathcal{M} \models \phi).$$

We denote the collection of all models of $\mathcal{T}$ by

$$\mathrm{Mod}(\mathcal{T}) := \{\mathcal{M} \mid \mathcal{M} \models \mathcal{T}\} = {}^{\models}\mathcal{T}.$$

As in propositional logic, this is the dual ${}^{\models}\mathcal{T}$ with respect to the Galois connection induced by the binary relation $\models$ between $\mathcal{A}$-structures and $\mathcal{A}$-sentences. Conversely, for a class of $\mathcal{A}$-structures $\mathcal{K}$,

$$\mathrm{Th}(\mathcal{K}) := \left\{\phi \in \mathcal{L}^{\varnothing}_{\mathsf{form}}(\mathcal{A}) \;\middle|\; \forall \mathcal{M} \in \mathcal{K}\, (\mathcal{M} \models \phi)\right\} = \mathcal{K}^{\models}.$$

A class $\mathcal{K}$ is **axiomatizable** if it is $\mathrm{Mod}(\mathcal{T})$ for some $\mathcal{T}$, in which case in fact $\mathcal{K} = \mathrm{Mod}(\mathrm{Th}(\mathcal{K}))$. For a theory $\mathcal{T}$, the sentences $\phi \in \mathrm{Th}(\mathrm{Mod}(\mathcal{T}))$ are the **semantic consequences** of $\mathcal{T}$.

2.22. **Exercise** (for set theorists)**.** Show that, in all but the most degenerate cases (which?), $\mathrm{Mod}(\mathcal{T})$ is a proper class, not a set.

2.23. **Example.** The **theory of (simple undirected) graphs** is the $\mathcal{A}_{\mathsf{graph}} = \{E\}$-theory

$$\mathcal{T}_{\mathsf{graph}} := \{\forall x\, \neg E(x,x),\ \forall x\, \forall y\, (E(x,y) \to E(y,x))\}.$$

An $\mathcal{A}_{\mathsf{graph}}$-structure $\mathcal{M} = (M, E^{\mathcal{M}})$, where $E^{\mathcal{M}} \subseteq M^2$, is a model of $\mathcal{T}_{\mathsf{graph}}$ iff

$$\forall a \in M\, ((a,a) \notin E^{\mathcal{M}}), \qquad \forall a, b \in M\, ((a,b) \in E^{\mathcal{M}} \implies (b,a) \in E^{\mathcal{M}}).$$

2.24. **Example.** The **theory of posets (partially ordered sets)** is the $\mathcal{A}_{\mathsf{poset}} = \{\leq\}$-theory

$$\begin{aligned}\mathcal{T}_{\mathsf{poset}} := \{&\forall x\, (x \leq x),\\ &\forall x\, \forall y\, \forall z\, ((x \leq y) \wedge (y \leq z) \to (x \leq z)),\\ &\forall x\, \forall y\, ((x \leq y) \wedge (y \leq x) \to (x = y))\}.\end{aligned}$$

An $\mathcal{A}_{\mathsf{poset}}$-structure $\mathcal{M} = (M, \leq^{\mathcal{M}})$ is a model of $\mathcal{T}_{\mathsf{poset}}$ iff $\leq^{\mathcal{M}}$ is a reflexive, transitive, and antisymmetric binary relation on M.

2.25. **Example.** The **theory of totally** or **linearly ordered sets** is the $\mathcal{A}_{\mathsf{poset}}$-theory

$$\mathcal{T}_{\mathsf{toset}} := \mathcal{T}_{\mathsf{poset}} \cup \{\forall x\, \forall y\, ((x \leq y) \vee (y \leq x))\}.$$

2.26. **Example.** The **theory of equivalence relations** is the $\mathcal{A}_{\mathsf{equiv}} = \{\sim\}$-theory consisting of the first two axioms of $\mathcal{T}_{\mathsf{poset}}$ together with the last axiom ("symmetry") of $\mathcal{T}_{\mathsf{graph}}$, with the relation symbols replaced by $\sim$.

As these examples show, it is often useful to "modularize" theories into groups of related axioms.

2.27. **Example.** The **theory of monoids** is the theory over the signature $\mathcal{A}_{\mathsf{mon}} = \{\cdot, 1\}$, where $\cdot, 1$ are binary, nullary function symbols respectively, with the axioms

$$\begin{aligned}\mathcal{T}_{\mathsf{mon}} := \{&\forall x\, \forall y\, \forall z\, ((x \cdot y) \cdot z = x \cdot (y \cdot z)),\\ &\forall x\, (1 \cdot x = x),\\ &\forall x\, (x \cdot 1 = x)\}.\end{aligned}$$

Examples of monoids are $(\mathbb{R}, +, 0)$, $(\mathbb{R}, \cdot, 1)$, $(X^X, \circ, \mathrm{id}_X)$, and $(\mathcal{P}(X), \cup, \varnothing)$ for any set X.

2.28. **Example.** The **theory of groups** over $\mathcal{A}_{\mathsf{grp}} = \mathcal{A}_{\mathsf{mon}} \cup \{^{-1}\}$, where $^{-1}$ is a unary function symbol (usually written after its argument), is

$$\mathcal{T}_{\mathsf{grp}} := \mathcal{T}_{\mathsf{mon}} \cup \{\forall x\,(x \cdot x^{-1} = 1), \forall x\,(x^{-1} \cdot x = 1)\}.$$

Examples are $(\mathbb{R}, +, 0, -)$, $((0,\infty), \cdot, 1, ^{-1})$, and the subset of X^X consisting of the bijections.

2.29. **Remark.** In many textbooks and courses, a monoid/group would be defined as a set equipped with merely an associative binary operation (called a **semigroup**), obeying the *axiom* that there exists an identity element and/or inverses. In other words, we may *define* the operation of identity/inverse from the primitive operation of multiplication; see Definition 2.64 below for the general development of this idea. However, having these additional definable operations as part of the signature is more convenient, especially in the case of monoids; see e.g., Exercise 2.47(b).

2.30. **Example.** An **abelian group** is a group which moreover satisfies the commutativity axiom. Usually, when discussing a generic abelian group, one uses an additive notation; we thus take $\mathcal{A}_{\mathsf{abgrp}} := \{+, 0, -\}$, and $\mathcal{T}_{\mathsf{abgrp}}$ to be the above theory $\mathcal{T}_{\mathsf{grp}}$ with $\cdot, 1, ^{-1}$ replaced with $+, 0, -$ respectively, plus the axiom

$$\forall x\,\forall y\,(x + y = y + x).$$

Models include $\mathbb{Z}$, $2\mathbb{Z} = \{2n \mid n \in \mathbb{Z}\}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{Z}/2\mathbb{Z}$, ... with the usual operations.

(An example of a *non-abelian* group is the symmetric group S_X of all bijections $X \cong X$ of a set X with ≥ 3 elements: swapping $a \leftrightarrow b$ does not commute with swapping $b \leftrightarrow c$.)

2.31. **Example.** The **theory of rings** is the $\mathcal{A}_{\mathsf{ring}} = \mathcal{A}_{\mathsf{field}} = \{+, 0, -, \cdot, 1\}$-theory $\mathcal{T}_{\mathsf{ring}}$ consisting of $\mathcal{T}_{\mathsf{abgrp}} \cup \mathcal{T}_{\mathsf{mon}}$ together with the distributivity axioms

$$\forall x\,\forall y\,\forall z\,(x \cdot (y + z) = x \cdot y + x \cdot z),$$
$$\forall x\,\forall y\,\forall z\,((y + z) \cdot x = y \cdot x + z \cdot x).$$

Both distributivity axioms are needed, because $\cdot$ might not be commutative; the **theory of commutative rings** is the $\mathcal{A}_{\mathsf{ring}}$-theory

$$\mathcal{T}_{\mathsf{commring}} := \mathcal{T}_{\mathsf{ring}} \cup \{\forall x\,\forall y\,(x \cdot y = y \cdot x)\}.$$

Examples of rings include $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/n\mathbb{Z}, \mathrm{Mat}_{n\times n}(\mathbb{R}) = \{n \times n \text{ matrices with real entries}\}$; all of these except the last are commutative.

(Some people don't require rings to have a multiplicative identity 1, in which case $n\mathbb{Z} \subseteq \mathbb{Z}$ would also be a ring. Those who require rings to contain 1 call $n\mathbb{Z}$ a **rng**, as in r*i*ng without *i*dentity.)

The **theory of fields** is the $\mathcal{A}_{\mathsf{field}}$-theory

$$\mathcal{T}_{\mathsf{field}} := \mathcal{T}_{\mathsf{commring}} \cup \{\neg(0 = 1),\ \forall x\,(\neg(x = 0) \to \exists y\,(x \cdot y = 1))\}.$$

Models include $\mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/n\mathbb{Z}$ for prime n (only).

The **theory of ordered fields** is the $\mathcal{A}_{\mathsf{ordfield}}$-theory

$$\mathcal{T}_{\mathsf{ordfield}} := \mathcal{T}_{\mathsf{field}} \cup \mathcal{T}_{\mathsf{toset}} \cup \{\forall x\,\forall y\,\forall z\,((x \leq y) \to (x + z \leq y + z)),\ \forall x\,\forall y\,\forall z\,((x \leq y) \wedge (0 \leq z) \to (x \cdot z \leq y \cdot z))\}.$$

Models include $\mathbb{Q}, \mathbb{R}$ (not $\mathbb{C}$ or $\mathbb{Z}/n\mathbb{Z}$).

2.32. **Exercise.** One would perhaps think that $\cdot$ in a ring should be required to distribute over not just binary $+$, but also nullary 0 and perhaps "(-1)-ary" negation $-$.

Show that this is (semantically) implied by distributivity over $+$ along with the other ring axioms, but would *not* be implied if we only required addition to form a monoid, not a group. (See Exercises 2.47 and 3.29. A ring without $-$ is sometimes called a **rig**, as in ri*n*g without *n*egation; an example is given by $(\mathcal{P}(X), \cup, \varnothing, \cap, X)$.)

2.33. **Example.** The **theory of ($\mathbb{R}$-)vector spaces** is the *uncountable* $\mathcal{A}_{\mathsf{vec}}$-theory

$$\mathcal{T}_{\mathsf{vec}} := \mathcal{T}_{\mathsf{abgrp}} \cup \left\{ \begin{array}{l} \forall x\,((ab)\cdot x = a\cdot(b\cdot x)) \\ \forall x\,(1\cdot x = x) \\ \forall x\,((a+b)\cdot x = a\cdot x + b\cdot x) \\ \forall x\,\forall y\,(a\cdot(x+y) = a\cdot x + a\cdot y) \end{array} \middle| \; a,b\in\mathbb{R} \right\}.$$

Note that $\forall x$ quantifies over elements of the structure (i.e., vectors), while to express laws which hold for each scalar, we need infinite families of axioms, one for each scalar. (Note also the differing roles of a, b, x in e.g., the third axiom, in which the terms on either side have the tree structures

In particular, note that $+$ in $a+b$ does *not* refer to the symbol $+\in\mathcal{A}_{\mathsf{vec}}$, but to addition in $\mathbb{R}$.)

2.34. **Example.** Let G be a monoid. An **(left) action** of G on a set X is a function $\cdot : G\times X\to X$ obeying only the first two axioms above, which are the only ones that make sense:

$$\forall x\,((ab)\cdot x = a\cdot(b\cdot x)),$$
$$\forall x\,(1\cdot x = x).$$

A **G-set** is a set equipped with an action of G. The **theory $\mathcal{T}_{G\mathsf{set}}$ of (left) G-sets** (for a fixed G) consists of these two axioms, over the signature $\mathcal{A}_{G\mathsf{set}} = \{a\cdot \mid a\in G\}$.

(Thus, a $\mathbb{R}$-vector space, or more generally K-vector space for any field K, is a K-set, where K refers to the multiplicative monoid, such that the action $K\times X\to X$ preserves $+$ on both sides.)

2.35. **Example.** Suppose we wish to axiomatize structures which are sets M equipped with an injective sequence $f:\mathbb{N}\to M$ of distinct elements. Since $n\in\mathbb{N}$ is not in M, we cannot treat f as a unary function; rather, for each $n\in\mathbb{N}$, we treat $f(n)$ as a *single constant symbol*. Let

$$\mathcal{A} := \{f(0), f(1), f(2), \dots\}.$$

An $\mathcal{A}$-structure $\mathcal{M}$ is a set M equipped with $f(0)^{\mathcal{M}}, f(1)^{\mathcal{M}}, \dots \in M$. Now to impose injectivity, take

$$\mathcal{T} := \{\neg(f(m) = f(n)) \mid m\neq n\in\mathbb{N}\}.$$

(Note that it is not possible to enforce *surjectivity* of f via a first-order theory. Intuitively, this is because we would need to say $\forall x\,((x=f(0))\vee(x=f(1))\vee\cdots)$; see Theorem 4.3.)

Finally, here are some degenerate examples of theories:

2.36. **Example.** The $\varnothing$-theory $\varnothing$ axiomatizes the class of all sets (i.e., $\varnothing$-structures).

2.37. **Example.** The $\varnothing$-theory $\mathcal{T} = \{\forall x\,\forall y\,(x=y)\}$ axiomatizes the class of sets of size ≤ 1.

2.38. **Exercise.** Verify that a finite union of axiomatizable classes of structures is still axiomatizable.

2.39. **Exercise.** Show that for any finite $F\subseteq\mathbb{N}$, the class of all sets of size $\in F$ is axiomatizable.

2.40. **Remark.** For a theory $\mathcal{T}$ with free variables from X, a "model of $\mathcal{T}$" cannot just be a structure; we also need to know how the variables are assigned, before each $\phi\in\mathcal{T}$ has a definite truth value. Thus, we take a **model of $\mathcal{T}$** to mean a pair $(\mathcal{M},\alpha)$ of a structure $\mathcal{M}$ together with a variable assignment $\alpha: X\to M$ such that $\mathcal{M}\models_\alpha\phi$, and denote the collection of all such by

$$\mathrm{Mod}_X(\mathcal{T}) := \{(\mathcal{M},\alpha) \mid \alpha: X\to M \text{ and } \mathcal{M}\models_\alpha\mathcal{T}\}.$$

Everything we said above still holds for this extended notion of model (for fixed X).

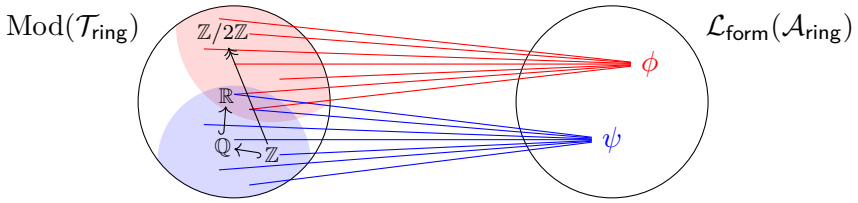
2.C. **Homomorphisms and definability.** These are an entirely new feature of first-order logic. Whereas two propositional models $m, n : \mathcal{A} \to \{0, 1\}$ can only be equal or not, in first-order logic, two models $\mathcal{M}, \mathcal{N}$ may look "the same" for all intents and purposes, without actually being equal:

2.41. **Example.** There are two possible ways of defining the ring (or just abelian group) $\mathbb{Z}/n\mathbb{Z}$: concretely, as the numbers $0, 1, \dots, n-1$ with arithmetic operations followed by taking remainder; or abstractly, as congruence classes $[0], [1], \dots, [n-1]$ mod n. These two are technically not the same structure (the elements of the former are natural numbers, whereas the elements of the latter are equivalence classes of integers); rather, they are two "equivalent copies of the same structure".

2.42. **Example.** Since $\mathbb{R}$ is constructed from $\mathbb{Q}$ (e.g., as Dedekind cuts, or equivalence classes of Cauchy sequences), we technically do not have $\mathbb{Q} \subseteq \mathbb{R}$; rather, there is a "copy" of $\mathbb{Q}$ inside $\mathbb{R}$.

The big picture is as follows. As before, we have the left "semantic side" of first-order logic, consisting of models, as well as the right "syntactic side", consisting of formulas; and these are related by the Galois connection induced by the satisfaction relation $\models$.

(2.43)



The difference is that now, the elements of the left side of the picture (i.e., models) may be compared and related amongst themselves: for example, the ring $\mathbb{Q}$ may be "embedded" inside $\mathbb{R}$, while $\mathbb{Z}$ has $\mathbb{Z}/2\mathbb{Z}$ as a quotient. These relationships among models, or more precisely, "structure-preserving maps", in turn interact with the interpretation of the formulas on the right side in various ways.

2.44. **Definition.** Let $\mathcal{A}$ be a signature, $\mathcal{M}, \mathcal{N}$ be two $\mathcal{A}$-structures. An **$\mathcal{A}$-homomorphism** $h : \mathcal{M} \to \mathcal{N}$ is a function $h : M \to N$ between their underlying sets which preserves the structure:

(a) for each n-ary function symbol $f \in \mathcal{A}^n_{\mathsf{fun}}$, we have
$$h(f^{\mathcal{M}}(a_1, \dots, a_n)) = f^{\mathcal{N}}(h(a_1), \dots, h(a_n))$$
for all $a_1, \dots, a_n \in M$ (we express this as "h **preserves (the interpretation of)** f");

(b) for each n-ary relation symbol $R \in \mathcal{A}^n_{\mathsf{rel}}$, we have the equivalent conditions
$$(a_1, \dots, a_n) \in R^{\mathcal{M}} \implies (h(a_1), \dots, h(a_n)) \in R^{\mathcal{N}},$$
$$R^{\mathcal{M}}(a_1, \dots, a_n) \leq R^{\mathcal{N}}(h(a_1), \dots, h(a_n)).$$
Note the $\leq$, *not* $=$! (We express this as "h **preserves (the interpretation of)** R".)

We let
$$\mathrm{Hom}(\mathcal{M}, \mathcal{N}) := \big\{ h \in N^M \;\big|\; h \text{ is a homomorphism} \big\}.$$

2.45. **Example.** For two $\mathcal{A}_{\mathsf{abgrp}}$-structures $\mathcal{M}, \mathcal{N}$, a homomorphism $h : \mathcal{M} \to \mathcal{N}$ has to obey

$(*)$
$$h(a +^{\mathcal{M}} b) = h(a) +^{\mathcal{N}} h(b),$$
$$h(0^{\mathcal{M}}) = 0^{\mathcal{N}},$$
$$h(-^{\mathcal{M}} a) = -^{\mathcal{N}} h(a).$$

Note that these conditions have nothing to do with what axioms $\mathcal{M}, \mathcal{N}$ *satisfy*! If $\mathcal{M}, \mathcal{N}$ happen to be abelian groups (i.e., models of $\mathcal{T}_{\mathsf{abgrp}}$), then we call h an **abelian group homomorphism**. For example, $\exp : (\mathbb{R}, +, 0, -) \to (\mathbb{R}, \cdot, 1, {}^{-1})$ is a homomorphism, by the exponent laws: $(*)$ says
$$e^{a+b} = e^a \cdot e^b.$$

2.46. **Example.** For two vector spaces $\mathcal{M}, \mathcal{N} \models \mathcal{T}_{\mathsf{vec}}$, a homomorphism $h : \mathcal{M} \to \mathcal{N}$ has to obey the above, as well as, for each $r \in \mathbb{R}$ and $a \in M$,

$$h(r \cdot^{\mathcal{M}} a) = r \cdot^{\mathcal{N}} h(a). \tag{$\dagger$}$$

Of course, homomorphisms of vector spaces are usually called **linear transformations**.

Linear transformations are usually defined (in a linear algebra class, say) by requiring only the two conditions $(*)$ and $(\dagger)$. This is an accident specific to certain kinds of structures, where preservation of some parts of the structure implies preservation of others; the general notion of homomorphism, which works for all structures, requires preservation of everything.

2.47. **Exercise.**

(a) Show that if a function $h : \mathcal{M} \to \mathcal{N}$ between groups preserves $\cdot$, then it also preserves $1, {}^{-1}$, i.e., it is a group homomorphism. [See also Exercise 2.62.]
(b) Thus, if $\mathcal{M}, \mathcal{N}$ are vector spaces and h preserves $+, r\cdot$, then h is a linear transformation.
(c) Thus, in a ring (Example 2.31), $\cdot$ distributes over (not only $+$ but also) 0 and $-$.
(d) Give an example of a non-homomorphism between monoids which nonetheless preserves $\cdot$.

2.48. **Example.** For two posets $\mathcal{M}, \mathcal{N} \models \mathcal{T}_{\mathsf{poset}}$, a homomorphism $h : \mathcal{M} \to \mathcal{N}$ has to obey

$$a \leq^{\mathcal{M}} b \implies h(a) \leq^{\mathcal{N}} h(b).$$

This is usually called an **order-preserving** or **monotone** function. For example, $\exp : \mathbb{R} \to \mathbb{R}$ is monotone. For a non-numeric example, for any function $f : X \to Y$ between sets, taking preimage yields a monotone function between their powersets:

$$f^{-1} : \mathcal{P}(Y) \longrightarrow \mathcal{P}(X).$$

This is not only order-preserving:

$$A \subseteq B \subseteq Y \implies f^{-1}(A) \subseteq f^{-1}(B),$$

but also preserves the set operations $\cap, \cup$, etc.:

$$\begin{aligned} f^{-1}(A \cap B) &= f^{-1}(A) \cap f^{-1}(B), \\ f^{-1}(A \cup B) &= f^{-1}(A) \cup f^{-1}(B), \\ f^{-1}(\varnothing) &= \varnothing), \\ f^{-1}(Y) &= X. \end{aligned}$$

(In other words, it is an *ordered rig homomorphism*; see Exercise 2.32.)

2.49. **Definition.** A homomorphism $h : \mathcal{M} \to \mathcal{N}$ is an **$\mathcal{A}$-isomorphism**, written $h : \mathcal{M} \cong \mathcal{N}$, if

(i) it is a bijection, i.e., it has an inverse $h^{-1} : N \to M$, which *a priori* is just a function;
(ii) h^{-1} is also a homomorphism $\mathcal{N} \to \mathcal{M}$.

Two structures $\mathcal{M}, \mathcal{N}$ are **isomorphic**, denoted $\mathcal{M} \cong \mathcal{N}$, if there exists an isomorphism $h : \mathcal{M} \cong \mathcal{N}$.

An **automorphism** of $\mathcal{M}$ is an isomorphism from $\mathcal{M}$ to itself. Let

$$\mathrm{Aut}(\mathcal{M}) := \{h \in \mathrm{Hom}(\mathcal{M}, \mathcal{M}) \mid h \text{ is an automorphism}\} \subseteq M^M.$$

2.50. **Example.** $\exp : (\mathbb{R}, +, 0, -) \to (\mathbb{R}, \cdot, 1, {}^{-1})$ is an isomorphism of abelian groups.

2.51. **Example.** It is a fundamental theorem of linear algebra that every finite-dimensional vector space (over $\mathbb{R}$) is isomorphic to $\mathbb{R}^n$ for some n.

In group theory and linear algebra, isomorphisms are often defined without mentioning condition 2.49(ii). Indeed, 2.49(ii) is automatic whenever the signature $\mathcal{A}$ in question is *functional*, i.e., consists only of function symbols; see Proposition 2.77. However, 2.49(ii) is needed in general:

2.52. **Example.** For two $\mathcal{A}$-structures $\mathcal{M}_1, \mathcal{M}_2$ on the same underlying set M, the identity function $M \to M$ is a homomorphism $\mathcal{M}_1 \to \mathcal{M}_2$ iff the function symbols $f \in \mathcal{A}_{\mathsf{fun}}$ are interpreted the same way, while for each relation symbol $R \in \mathcal{A}_{\mathsf{rel}}$, we have $R^{\mathcal{M}_1} \subseteq R^{\mathcal{M}_2}$.

Thus for example, $\mathrm{id}_{\mathbb{R}} : (\mathbb{R}, =) \to (\mathbb{R}, \leq)$ is a poset homomorphism, but not an isomorphism, since for the inverse $\mathrm{id}_{\mathbb{R}} : (\mathbb{R}, \leq) \to (\mathbb{R}, =)$ to be a homomorphism, we would need $(\leq) \subseteq (=)$.

A homomorphism by definition preserves function and relation symbols. We now show that they preserve most other concepts that can be defined from the primitive functions and relations, while isomorphisms preserve *everything*.

2.53. **Proposition.** Let $h : \mathcal{M} \to \mathcal{N}$ be a homomorphism.

(a) For each term $t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$ and variable assignment $\alpha : X \to M$, we have

$$h(t^{\mathcal{M}}(\alpha)) = t^{\mathcal{N}}(h \circ \alpha).$$

We express this by saying that "h **preserves (the interpretation of)** t".

(b) For each formula $\phi \in \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$ and variable assignment $\alpha : X \to M$, we have

$$\mathcal{M} \models_\alpha \phi \implies \mathcal{N} \models_{h\circ\alpha} \phi$$
$$\text{or equivalently} \quad \phi^{\mathcal{M}}(\alpha) \leq \phi^{\mathcal{N}}(h \circ \alpha),$$

provided that ϕ *is* **positive-existential**, i.e., built using only atomic formulas, $\wedge, \vee, \top, \bot$, and $\exists$ (without $\neg$, hence also without $\to$ or $\forall$).

We express this by saying that "h **preserves (the interpretation of)** ϕ".

(c) If h is an *iso*morphism, then h preserves the interpretations of all formulas. It follows that (by replacing ϕ with $\neg\phi$) the converse of the implication in (b) also holds.

Note that if we think of $\alpha \in M^X$ as an "X-ary tuple" $(\alpha(x))_{x\in X}$, then these are the analogues of the conditions in Definition 2.44 of homomorphism, extended from atomic to arbitrary formulas.

Proof. (a) By induction on t.

- For a variable $t = x \in X$, we have

$$\begin{aligned} h(x^{\mathcal{M}}(\alpha)) &= h(\alpha(x)) && \text{by definition of } x^{\mathcal{M}} \text{ (2.7)} \\ &= x^{\mathcal{N}}(h \circ \alpha) && \text{by definition of } x^{\mathcal{N}}. \end{aligned}$$

- For $t = f(t_1, \ldots, t_n)$ where $f \in \mathcal{A}^n_{\mathsf{fun}}$ and $t_1, \ldots, t_n \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$, we have

$$\begin{aligned} h(f(t_1, \ldots, t_n)^{\mathcal{M}}(\alpha)) &= h(f^{\mathcal{M}}(t_1^{\mathcal{M}}(\alpha), \ldots, t_n^{\mathcal{M}}(\alpha))) && \text{by definition of } f(t_1, \ldots, t_n)^{\mathcal{M}} \\ &= f^{\mathcal{N}}(h(t_1^{\mathcal{M}}(\alpha)), \ldots, h(t_n^{\mathcal{M}}(\alpha))) && \text{since } h \text{ is a homomorphism} \\ &= f^{\mathcal{N}}(t_1^{\mathcal{N}}(h \circ \alpha), \ldots, t_n^{\mathcal{N}}(h \circ \alpha)) && \text{by IH} \\ &= f(t_1, \ldots, t_n)^{\mathcal{N}}(h \circ \alpha) && \text{by definition of } f(t_1, \ldots, t_n)^{\mathcal{N}}. \end{aligned}$$

(b) By induction on ϕ. We will state the inductive cases as a series of lemmas, as they are sometimes also useful on their own. (See the examples after this proof.)

2.54. **Lemma.** Every homomorphism $h : \mathcal{M} \to \mathcal{N}$ preserves the interpretation of atomic formulas.

Proof. Similar to the inductive case of (a) (**Exercise**). □

2.55. **Lemma.** Let $h : M \to N$ be an arbitrary function (between the underlying sets).

(i) If h preserves $\phi, \psi \in \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$, then h also preserves $\phi \wedge \psi$, $\phi \vee \psi$.

(ii) h always preserves $\top, \bot$.

Proof. (i) The key point is that the functions $\min, \max : \{0,1\}^2 \to \{0,1\}$ used to interpret $\wedge, \vee$ are *monotone*. Thus for any variable assignment $\alpha : X \to M$, we have

$$\begin{aligned}(\phi \wedge \psi)^{\mathcal{M}}(\alpha) &= \min(\phi^{\mathcal{M}}(\alpha), \psi^{\mathcal{M}}(\alpha)) && \text{by definition}\\ &\leq \min(\phi^{\mathcal{N}}(h \circ \alpha), \psi^{\mathcal{N}}(h \circ \alpha)) && \text{by preservation of } \phi, \psi \text{ and monotonicity of min}\\ &= (\phi \wedge \psi)^{\mathcal{N}}(h \circ \alpha) && \text{by definition,}\end{aligned}$$

and similarly for $\vee$. (This doesn't work for $\neg$, because $x \mapsto 1 - x$ is not monotone.)

(ii) $\top^{\mathcal{M}}(\alpha) = 1 \leq 1 = \top^{\mathcal{N}}(h \circ \alpha)$; similarly for $\bot$. □

2.56. **Lemma.** Let $h : M \to N$ be an arbitrary function. If h preserves the interpretation of $\phi \in \mathcal{L}^{X \cup \{x\}}_{\mathsf{form}}(\mathcal{A})$, where x is any variable, then h also preserves the interpretation of $\exists x\, \phi \in \mathcal{L}^{X}_{\mathsf{form}}(\mathcal{A})$.

Proof. Let $\alpha : X \to M$. We have

$$\begin{aligned}\mathcal{M} \models_{\alpha} \exists x\, \phi &\iff \exists a \in M \text{ s.t. } \mathcal{M} \models_{\alpha\langle x \mapsto a\rangle} \phi,\\ (*) \qquad \mathcal{N} \models_{h \circ \alpha} \exists x\, \phi &\iff \exists b \in N \text{ s.t. } \mathcal{N} \models_{(h \circ \alpha)\langle x \mapsto b\rangle} \phi;\end{aligned}$$

and we must show that the former implies the latter. Given the former, since h preserves ϕ, we get

$$\mathcal{N} \models_{h \circ \alpha\langle x \mapsto a\rangle} \phi.$$

Now $h \circ \alpha\langle x \mapsto a\rangle = (h \circ \alpha)\langle x \mapsto h(a)\rangle : X \cup \{x\} \to N$, since both functions map $x \mapsto h(a)$ and all other $y \in X \setminus \{x\}$ to $h(\alpha(y))$. Thus the RHS of $(*)$ holds with $b := h(a)$. □

Now the proof of 2.53(b) follows immediately by induction on ϕ.

(c) Since h, h^{-1} are both homomorphisms, by Lemma 2.54, both preserve atomic formulas. But

2.57. **Lemma.** If $h : \mathcal{M} \to \mathcal{N}$ is a bijection, then h preserves $\phi \in \mathcal{L}^{X}_{\mathsf{form}}(\mathcal{A})$ iff h^{-1} preserves $\neg\phi$.

Proof. To show $\Longrightarrow$: $\mathcal{N} \models_{\alpha} \neg\phi \iff \mathcal{N} \not\models_{\alpha = h \circ h^{-1} \circ \alpha} \phi \implies \mathcal{M} \not\models_{h^{-1} \circ \alpha} \phi \iff \mathcal{M} \models_{h^{-1} \circ \alpha} \neg\phi$. □

Together with the preceding lemmas, we may now show by induction that h, h^{-1} both preserve arbitrary formulas. (We only care about h in the end, but need to also show it for h^{-1} in order for the inductive case for $\neg$ to go through.) (*Proof of* 2.53)□

2.58. **Example.** For a homomorphism $h : \mathcal{M} \to \mathcal{N}$ between vector spaces, for the $\mathcal{A}_{\mathsf{vec}}$-term

$$t := 3 \cdot x + 4 \cdot (5 \cdot y + (-z)),$$

Proposition 2.53(a) says that for all $a, b, c \in M$, (omitting superscripts $^{\mathcal{M}}$, $^{\mathcal{N}}$ for clarity)

$$h(3 \cdot a + 4 \cdot (5 \cdot b + (-c))) = 3 \cdot h(a) + 4 \cdot (5 \cdot h(b) + (-h(c))).$$

In other words, we recover the familiar fact that linear transformations preserve linear combinations.

2.59. **Example.** For a homomorphism $h : \mathcal{M} \to \mathcal{N}$ between commutative rings (see Example 2.31), we can likewise think of a term $t \in \mathcal{L}^{X}_{\mathsf{term}}(\mathcal{A}_{\mathsf{ring}})$ as a polynomial with integer coefficients, e.g.,

$$t = x \cdot x + x \cdot x + x \cdot x + (-x) + 1 + 1$$

would be how we formally write $3x^2 - x + 2$; h then has to preserve evaluation of all such polynomials:

$$h(3z^2 - a + 2) = h(t^{\mathcal{M}}(x \mapsto a)) = t^{\mathcal{N}}(h \circ (x \mapsto a)) = 3h(a)^2 - h(a) + 2.$$

If we now consider the positive-existential sentence

$$\phi := \exists x\, (t = 0),$$

then Proposition 2.53(b) says that if the polynomial t has a root in $\mathcal{M}$, then it also has a root in $\mathcal{N}$:

$$\exists a \in M \text{ s.t. } t^{\mathcal{M}}(x \mapsto a) = 0 \iff \mathcal{M} \models \phi \implies \mathcal{N} \models \phi \iff \exists b \in N \text{ s.t. } t^{\mathcal{N}}(x \mapsto b) = 0.$$

2.60. **Example.** Consider the $\mathcal{A}_{\mathsf{ring}}$-formula with one free variable x

$$\phi := \exists y\,(x \cdot y = 1).$$

The interpretation in a commutative ring $\mathcal{M}$ under an assignment $x \mapsto a$ says that a has a multiplicative inverse. Thus, *commutative ring homomorphisms preserve invertibility.*

In a field, we have

$$\begin{aligned} x \neq y &\iff x + (-y) \neq 0 \\ &\iff x + (-y) \text{ is invertible} \\ &\iff \exists z\,((x + (-y)) \cdot z = 1). \end{aligned}$$

Thus, *field homomorphisms preserve* $\neq$*, i.e., are injective.*

2.61. **Example.** Consider the formula

$$x \cdot x = x.$$

If a function h between groups preserves $\cdot$, then it preserves this formula; but this formula is equivalent in a group to "$1 = x$", whence h preserves 1.

2.62. **Exercise.** Similarly, show that a function between groups preserving $\cdot$ and 1 also preserves $^{-1}$, hence is a group homomorphism (reproving Exercise 2.47).

2.63. **Exercise.** Show in general that $h : \mathcal{M} \to \mathcal{N}$ preserves the interpretation of a term $t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$ iff it preserves the interpretation of the formula $t = y$, where y is an additional variable not in X.

These examples show that if a concept (operation or relation) can be expressed in a positive-existential way in a given structure, then it is in some sense "implicitly part of the structure". We formalize this idea as follows:

2.64. **Definition.** Let $\mathcal{K}$ be a class of structures, and suppose we have an assignment $\mathcal{M} \mapsto R_{\mathcal{M}}$ to each $\mathcal{M} \in \mathcal{K}$ of some n-ary relation $R_{\mathcal{M}} \subseteq M^n$ (not necessarily denoted by any relation symbol in the signature). We say that the family of these relations $R_{\mathcal{M}}$ is **definable** (in all structures in $\mathcal{K}$) if there is an $\mathcal{A}$-formula $\phi \in \mathcal{L}^{\{x_1,\dots,x_n\}}_{\mathsf{form}}(\mathcal{A})$ such that for every $\mathcal{M} \in \mathcal{K}$, $R_{\mathcal{M}}$ is given by

$$\phi^{\mathcal{M}}_{\{x_1,\dots,x_n\}} \subseteq M^{\{x_1,\dots,x_n\}} \cong M^n.$$

In other words,

$$(a_1, \dots, a_n) \in R_{\mathcal{M}} \iff \mathcal{M} \models_{x_1 \mapsto a_1, \dots, x_n \mapsto a_n} \phi.$$

If ϕ is defined by a special type of formula, e.g., positive-existential, then we say it is **positive-existentially definable**.

If we instead have an assignment $\mathcal{M} \mapsto f_{\mathcal{M}}$ of an n-ary function $f_{\mathcal{M}} : M^n \to M$ for each $\mathcal{M} \in \mathcal{K}$, then we say these functions are **definable** if their graphs are definable (as $(n+1)$-ary relations).

2.65. **Example.** By Example 2.60, $\neq$ is positive-existentially definable in fields. (It is also definable, but not positive-existentially definable, in *all* $\mathcal{A}_{\mathsf{field}}$-structures, by Corollary 2.67, since not every $\mathcal{A}_{\mathsf{field}}$-homomorphism is injective, e.g., between two rings which are not fields such as $\mathbb{Z} \twoheadrightarrow \mathbb{Z}/2\mathbb{Z}$.)

2.66. **Example.** By Example 2.61 and Exercise 2.62, the functions 1 (the constant) and $^{-1}$ are positive-existentially definable from $\{\cdot\}$ in groups.

1 is also definable in monoids: $x = 1 \iff \forall y\,((x \cdot y = y) \wedge (y \cdot x = y))$. However, it is not positive-existentially definable, by Exercise 2.47(b).

2.67. **Corollary** (of Proposition 2.53 and Exercise 2.63)**.** If a relation is definable in every structure in $\mathcal{K}$, then it is preserved by every isomorphism between two such structures, and by every homomorphism if it is positive-existentially definable.

Similarly for definable functions. □

2.68. **Example.** In the field $\mathbb{R}$ (so $\mathcal{K} = \{\mathbb{R}\}$), we have

$$0 \leq x \iff \exists y\,(y \cdot y = x),$$
$$x \leq y \iff 0 \leq y + (-x) \iff \exists z\,(z \cdot z = y + (-x)),$$

i.e., $\leq$ is positive-existentially definable. Thus, every field homomorphism $\mathbb{R} \to \mathbb{R}$ is monotone.

2.69. **Exercise.** Show that $\leq$ is positive-existentially definable in the monoid $\mathbb{N}$ under addition.

2.70. **Exercise.** Show that $\leq$ is positive-existentially definable in the field $\mathbb{Q}$.
[Hint: **Lagrange's four-squares theorem** says every $n \in \mathbb{N}$ is a sum of four perfect squares.]

2.71. **Example.** The constant $i \in \mathbb{C}$ is *not* definable from the field structure. That is, its graph, which is the singleton $\{i\} \subseteq \mathbb{C}$, is not definable: there is no formula $\phi \in \mathcal{L}^{\{x\}}_{\mathsf{form}}(\mathcal{A}_{\mathsf{field}})$ true only at i. This is because the complex conjugation map $z \mapsto \overline{z}$ is a field automorphism not fixing i.

This makes precise the oft-quoted idea that the imaginary unit i is "not uniquely defined", because $-i$ is also a square root of -1 and there's no way to tell them apart: if you took every mathematical document ever written and changed all the i's to $-i$'s, everything would remain valid. Of course, this all depends on the background notions (operations and relations) involving complex numbers that one is allowed to talk about. If we introduce a constant symbol denoting i, or something equivalent (like a unary relation "has positive imaginary part"), complex conjugation would no longer be an automorphism of the resulting structure!

2.72. **Example.** For any subset $N \subseteq M$ of the underlying set of a structure $\mathcal{M}$, closed under $f^{\mathcal{M}}$ for all $f \in \mathcal{A}_{\mathsf{fun}}$, we have a **substructure** $\mathcal{N}$ on N defined by restricting the interpretations in $\mathcal{M}$ of all the symbols in $\mathcal{A}$. The inclusion function $i : N \hookrightarrow M$ is always a homomorphism from such a substructure (with $\Longrightarrow$ strengthened to $\iff$ in the preservation of relations in Definition 2.44).

For example, $\mathbb{Z} \subseteq \mathbb{R}$ is a $\{0, 1, \leq\}$-substructure (under the usual interpretations). The sentence

$$\phi := \forall x\,((x \leq 0) \vee (1 \leq x))$$

is true in $\mathbb{Z}$, but not in $\mathbb{R}$, hence is not preserved by the inclusion $i : \mathbb{Z} \hookrightarrow \mathbb{R}$. Thus, by Proposition 2.53, it cannot be semantically equivalent in every ring to a positive-existential sentence.

2.73. **Exercise.** Show that inclusion of a substructure preserves all **existential** formulas, i.e., formulas built by starting with a quantifier-free formula, and then applying $\exists$ some number of times.

2.74. **Exercise.**

(a) Show that $\mathbb{Q}[\sqrt{2}] := \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\} \subseteq \mathbb{R}$ is a subring that is also a field.
(b) Show that $a + b\sqrt{2} \mapsto a - b\sqrt{2}$ is a well-defined automorphism of this field. [Hint: your argument must break if $\mathbb{Q}$ were replaced with $\mathbb{R}$!]
(c) Conclude that $\leq$ is *not* definable from the ring operations in $\mathbb{Q}[\sqrt{2}]$ (even though $\mathbb{Q}[\sqrt{2}]$ is sandwiched between the subfield $\mathbb{Q}$ and the superfield $\mathbb{R}$ in which $\leq$ *is* definable!).

2.75. **Exercise.** Show that any *surjective* homomorphism preserves all **positive** formulas, i.e., formulas built using $\wedge, \vee, \top, \bot, \exists, \forall$ (no $\neg$, except those included in $\forall := \neg\exists\neg$).

2.76. **Exercise.** Find an example of a surjective homomorphism that fails to preserve $\neg$.

We end with a more theoretical "meta-application" of definability:

2.77. **Proposition.** Let $\mathcal{A}$ be a **functional** signature, i.e., consisting entirely of function symbols. Then every bijective $\mathcal{A}$-homomorphism $h : \mathcal{M} \to \mathcal{N}$ is an isomorphism.

Proof. Let $f \in \mathcal{A}^n_{\mathsf{fun}}$. We need to show that h^{-1} preserves f. By Exercise 2.63, this is equivalent to h^{-1} preserving its graph $f(x_1, \ldots, x_n) = y$. By Lemma 2.57, this is equivalent to h preserving $\neg(f(x_1, \ldots, x_n) = y)$, which holds true because h preserves f and is injective (so preserves $\neq$). $\square$

2.D. **Optional: infinitary logic.** Very broadly speaking, the concept of *definability* introduced in the preceding subsection is the first-order analogue of *axiomatizability* in propositional logic: both ask whether a given semantic concept has a syntactic definition. However, above we discussed only *certain* aspects of first-order definability; and these aspects are in some sense completely orthogonal to the aspects of axiomatizability that we discussed back in propositional logic. Whereas propositional axiomatizability is all about the limited expressive power of *finite* formulas (which hence had to do with "limits of truth assignments", i.e., approximating on any finite subset), the syntactic aspects of definability considered above have nothing at all to do with finiteness.

2.78. **Example.** In a vector space V (over $\mathbb{R}$, say), vectors $\vec{v}_1, \dots, \vec{v}_n$ are linearly dependent iff

$$\exists a_1, \dots, a_n \in \mathbb{R}, \text{ not all zero, s.t. } a_1\vec{v}_1 + \cdots + a_n\vec{v}_n = 0.$$

This is *not* a positive-existential first-order condition in the usual sense: the $\exists$ is not over the underlying set of the structure V; rather, it is really an infinite (indeed uncountable) disjunction

$$\bigvee_{\vec{0} \neq (a_1,\dots,a_n) \in \mathbb{R}^n} (a_1\vec{v}_1 + \cdots + a_n\vec{v}_n = 0).$$

Note that this formula, despite being infinite, is still "positive-existential" in an obvious sense.

Glancing at the proof of Proposition 2.53 should convince you that finiteness of the connectives is not used in any essential way. Moreover, one can easily think of many other common mathematical properties which can only be expressed in an infinitary way, to which it would be useful to apply the formalism of homomorphisms and isomorphisms from the preceding subsection (see below). The following exercises should give you a taste of such infinitary extensions of first-order logic:

2.79. **Exercise** (advanced)**. Infinitary first-order logic** (sometimes called $\mathcal{L}_{\infty\infty}$[3]) is defined as follows. An **infinitary signature** $\mathcal{A}$ is a set of function and relation symbols, where the arity of each symbol is an arbitrary set X;[4] as in first-order logic, we write $\mathcal{A}^X_{\mathsf{fun}}, \mathcal{A}^X_{\mathsf{rel}}$ to denote the X-ary function and relation symbols. The **infinitary $\mathcal{A}$-terms** are constructed inductively as follows:

- Every variable x is an $\mathcal{A}$-term.
- If $f \in \mathcal{A}^X_{\mathsf{fun}}$ is an X-ary function symbol, and for each $x \in X$, we have an $\mathcal{A}$-term $\tau(x)$ (so τ is a function from X to the collection of $\mathcal{A}$-terms), then $f(\tau)$ is an $\mathcal{A}$-term.

The **infinitary $\mathcal{A}$-formulas** are constructed as follows:

- If $R \in \mathcal{A}^X_{\mathsf{rel}}$ is an X-ary relation symbol, or the symbol $=$ when $X = \{0, 1\}$, and for each $x \in X$, we have an $\mathcal{A}$-term $\tau(x)$, then $R(\tau)$ is an **atomic $\mathcal{A}$-formula**.
- If $(\phi_i)_{i\in I}$ is an arbitrary family of $\mathcal{A}$-formulas indexed over some set I, then $\bigwedge_{i\in I} \phi_i, \bigvee_{i\in I} \phi_i$ are $\mathcal{A}$-formulas. When $I = \{0, 1\}$, we abbreviate $\bigwedge_{i\in I} \phi_i$ as $\phi_0 \wedge \phi_1$; when $I = \varnothing$, we abbreviate as $\top$. Similarly for $\vee, \bot$.
- If ϕ is an $\mathcal{A}$-formula, then so is $\neg\phi$.
- If ϕ is an $\mathcal{A}$-formula, and X is an arbitrary set of variables, then $\exists X\, \phi$ is an $\mathcal{A}$-formula.

The **free variables** of an infinitary term/formula are defined the usual way; the only subtle case is

$$\mathrm{FV}(\exists X\, \phi) := \mathrm{FV}(\phi) \setminus X.$$

Let as usual $\mathcal{L}^X_{\mathsf{term}}(\mathcal{A}), \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$ denote the terms/formulas with free variables from X.

(a) Define the notion of $\mathcal{A}$-structure. [For instance, you should be able to turn $\mathbb{R}$ into an $\mathcal{A} = \{L\}$-structure, where L is an $\mathbb{N}$-ary relation symbol whose interpretation says that a given sequence $\alpha : \mathbb{N} \to \mathbb{R}$ has a limit.]

[3]The first ∞ refers to the allowed arity of the $\bigwedge, \bigvee$'s; the second ∞ refers to the allowed arity of the $\exists$'s.

[4]If you prefer, you can assume that the arity is always an ordinal or even a cardinal, if you know what those are.

(b) Define the interpretation of $\mathcal{A}$-terms and $\mathcal{A}$-formulas in an $\mathcal{A}$-structure $\mathcal{M}$. [The $\exists$ case should be: for $\exists X\, \phi \in \mathcal{L}^Y_{\mathsf{form}}(\mathcal{A})$,

$$\mathcal{M} \models_\alpha \exists X\, \phi \;:\Longleftrightarrow\; \exists \beta : X \to M \text{ s.t. } \mathcal{M} \models_{\alpha\langle\beta\rangle} \phi,$$

where $\alpha\langle\beta\rangle$ is something you have to define.]

(c) Check that Proposition 2.53 goes through for infinitary formulas.

(d) Write down an infinitary positive-existential formula ϕ with free variables $x, y_1, \ldots, y_n$, in the *finitary* signature $\mathcal{A}_{\mathsf{vec}}$, whose interpretation in a vector space says that x is a linear combination of $y_1, \ldots, y_n$.

(e) Write down an infinitary positive-existential $\mathcal{A}_{\mathsf{vec}}$-formula with free variables $x_1, \ldots, x_n$ which says that $x_1, \ldots, x_n$ are linearly dependent.

(f) Write down an infinitary $\mathcal{A}_{\mathsf{vec}}$-sentence which says that a vector space has dimension n. Can the sentence be positive-existential?

(g) Write down an infinitary $\mathcal{A}_{\mathsf{poset}}$-sentence which says that every bounded increasing sequence has a least upper bound.

(h) An ordered field is **Archimedean** if every element in it is $\leq \overbrace{1 + \cdots + 1}^{n}$ for some $n \in \mathbb{N}$. Write down an infinitary $\mathcal{A}_{\mathsf{ordfield}}$-sentence which expresses this.

(i) A (simple undirected) graph is a **forest**, also called **acyclic**, if there are no loops (of length ≥ 3). Write down an infinitary $\mathcal{A}_{\mathsf{graph}}$-sentence which expresses this.

(j) Write down an infinitary $\varnothing$-sentence axiomatizing the finite sets.

(k) Find a suitable signature $\mathcal{A}$ and infinitary $\mathcal{A}$-sentence axiomatizing metric spaces.

2.80. **Exercise** (advanced)**.**

(a) Show that every rational in $\mathbb{R}$ is definable (as a constant) from the field structure.

(b) Show that *every* element of $\mathbb{R}$ is definable in infinitary logic.
[Hint: recall that $\leq$ is definable from the field structure, by Example 2.68.]

(c) Conclude that the only field automorphism $h : \mathbb{R} \cong \mathbb{R}$ is the identity function.

(d) Show that *every* subset $A \subseteq \mathbb{R}$ is infinitary definable from the field structure.

(e) Show that *every* n-ary relation $S \subseteq \mathbb{R}^n$ is infinitary definable from the field structure.

One benefit of infinitary logic is that we get exact converses to Corollary 2.67:

2.81. **Exercise** (advanced)**.** Let $\mathcal{A}$ be a (possibly infinitary) signature, $\mathcal{M}$ be an $\mathcal{A}$-structure.

(a) For any set X, give an infinitary formula $\phi \in \mathcal{L}^X_{\mathsf{form}}(\varnothing)$ such that for any $\alpha : X \to M$,

$$\mathcal{M} \models_\alpha \phi \iff \alpha : X \to M \text{ is a bijection.}$$

(b) Give an infinitary formula $\phi \in \mathcal{L}^M_{\mathsf{form}}(\mathcal{A})$ such that for any $\alpha : M \to M$,

$$\mathcal{M} \models_\alpha \phi \iff \alpha : \mathcal{M} \to \mathcal{M} \text{ is an automorphism.}$$

(c) For any $\vec{a} = (a_1, \ldots, a_n) \in M^n$, give an infinitary formula defining

$$\{g(\vec{a}) \mid g \in \mathrm{Aut}(\mathcal{M})\} \subseteq M^n.$$

[Hint: the formula should begin with $\exists M$, where here M is treated as a set of variables.]

(d) Conclude that if $R \subseteq M^n$ is preserved by every automorphism of $\mathcal{M}$, then R is definable by an infinitary formula.

(e) Prove that similarly, if $R \subseteq M^n$ is preserved by every homomorphism $\mathcal{M} \to \mathcal{M}$, then R is definable by an infinitary positive-existential formula.

2.82. **Remark.** It is a much deeper question, beyond the scope of this course, to ask precisely which relations $R \subseteq M^n$ are definable by formulas of bounded size, e.g., by finitary formulas, or by infinitary formulas which can use infinite $\bigwedge, \bigvee$ but only finite $\exists$.

2.E. **Application: affine geometry of the plane.** We now give an extended example of using the machinery of definability and homomorphisms to prove a nontrivial theorem in "normal" math.

We are interested in the geometry of the plane, $\mathbb{R}^2$. There are many concepts commonly thought of as "geometric", e.g., lines, circles, distances, angles, areas, etc. In 1872, Klein proposed the *Erlangen program*, an organizing framework that clarifies what "geometry" might mean:

- There is not a single "geometry", but rather multiple different "geometries", depending on which basic notions one is allowed to talk about.
- Each geometry is characterized by its allowed "symmetries". For example, the symmetries of Euclidean geometry include rotation, reflection, and translation.
- The meaningful notions in a particular geometry should be precisely those which are preserved by all the symmetries. For example, the symmetries of Euclidean geometry preserve lines and angles, but do not preserve, say, the notion of *vertical* line (which can be rotated), which is therefore not a meaningful notion in Euclidean geometry.

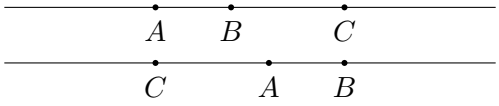
Of course, 1872 was before the advent of mathematical logic. In modern terminology, the various possible planar "geometries" are given by different structures (first-order or otherwise) one can put on the underlying set $\mathbb{R}^2$; while the "symmetries" of each "geometry" are the automorphisms of that structure. The last bullet point above is then an instance of the general fact that the notions definable in a structure are those preserved by all automorphisms (see Exercise 2.81).

2.83. **Definition. Affine geometry** is the geometry one gets by regarding lines as the only primitive notion. This is a more primitive geometry than Euclidean geometry: for example, scaling by a constant factor is an affine automorphism, which shows that "distance" is not an affine notion.

We can encode "lines" into a first-order structure by talking about collinearity of three points. Define the ternary relation on $\mathbb{R}^2$

$$\mathrm{Coll}(A, B, C) \iff A, B, C \text{ are collinear.}$$

Note that here, each of A, B, C is an ordered pair of real numbers. Note also that we do not necessarily require A, B, C to be distinct; if two of the points (or even all three) are equal, then they are trivially collinear. Note, finally, that we do not say anything about the *order* in which A, B, C appear on a line; for example, the following triples (A, B, C) both satisfy Coll:

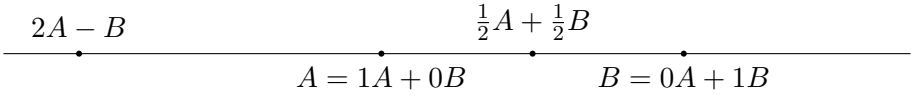


The following more quantitative notion allows us to talk about not just the ordering of points on a line, but even the precise position where a point appears on a line through two other points.

2.84. **Definition.** An **affine combination** of two points $A = (x_A, y_A)$ and $B = (x_B, y_B)$ is a point which can be written as a weighted (coordinatewise) average of them:

$$C = (1 - t)A + tB, \quad \text{for some } t \in \mathbb{R}.$$

In other words, an affine combination is just a linear combination where the coefficients sum to 1. For example, as t varies from 0 to 1, we move "at constant speed" from A to B.



Note that we also allow t outside the interval $[0, 1]$, in which case we get points on the line AB but not the line segment; e.g., when $t = -1$, we get $2A - B$, which can be written more intelligibly as $A + (A - B)$ (start at A, then shift by the vector from B to A). Note, finally, that we again do not require $A \neq B$; if $A = B$, then all affine combinations of them are the same.

In order to express affine combinations in first-order logic, we define the binary operation

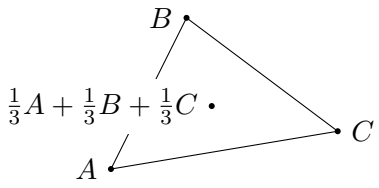
$$\begin{aligned} \mathrm{AC}_t : (\mathbb{R}^2)^2 &\longrightarrow \mathbb{R}^2 \\ (A, B) &\longmapsto (1-t)A + tB \end{aligned}$$

for each $t \in \mathbb{R}$. Thus we get an uncountable signature $\mathcal{A}_{\mathsf{aff}}$, consisting of the binary function symbols AC_t for each t (similarly to $\mathcal{A}_{\mathsf{vec}}$ from Example 1.7), together with a natural $\mathcal{A}_{\mathsf{aff}}$-structure on $\mathbb{R}^2$.

2.85. **Exercise.** More generally, an **affine combination** of any finite number of points $A_1, \dots, A_n \in \mathbb{R}^2$ is a point of the form

$$a_1A_1 + \cdots + a_nA_n \quad \text{for some } a_1, \dots, a_n \in \mathbb{R} \text{ adding to 1.}$$

For example, when $n = 3$ and $a_1 = a_2 = a_3 = \frac{1}{3}$, we get the **centroid** or **barycenter** of a triangle:



Show that for each n and $a_1, \dots, a_n$ adding to 1, the n-ary operation

$$\begin{aligned} \mathrm{AC}_{a_1,\dots,a_n} : (\mathbb{R}^2)^n &\longrightarrow \mathbb{R}^2 \\ (A_1, \dots, A_n) &\longmapsto a_1A_1 + \cdots + a_nA_n \end{aligned}$$

is definable from the $\mathcal{A}_{\mathsf{aff}}$-structure on $\mathbb{R}^2$. [Hint: when $n \geq 2$, not all the a_i can be 1.]

2.86. **Exercise.** Show that for any three points $A, B, C \in \mathbb{R}^2$, the following are equivalent:

(a) every point can be written as an affine combination of A, B, C in at least one way;
(b) every point can be written as an affine combination of A, B, C in at most one way;
(c) A, B, C are not collinear, i.e., none of them is an affine combination of the other two.

If these hold, we say A, B, C are an **affine basis** for $\mathbb{R}^2$.

We now have two reasonable candidates for how to formalize "affine geometry" as a first-order structure on $\mathbb{R}^2$: via either a single ternary relation $\{\mathrm{Coll}\}$, or all of the binary operations $\mathcal{A}_{\mathsf{aff}} = \{\mathrm{AC}_t\}_{t\in\mathbb{R}}$. It is intuitively clear that the latter is richer than merely the former: if we can say "C appears on the line AB at position t", for each t, then we can say that C appears at *some* position along the line. Formally, this "some" would be an infinite disjunction over all $t \in \mathbb{R}$. We also have to be slightly careful about the degenerate case where $A = B$; as noted above, in this case their only affine combination will be the same point $A = B$, even though any C will be trivially collinear with A, B, C (we can take the line AC instead). In summary, the infinitary formula

$$(A = B) \vee \bigvee_{t\in\mathbb{R}} (\mathrm{AC}_t(A, B) = C) \in \mathcal{L}_{\mathsf{form}}^{\{A,B,C\}}(\mathcal{A}_{\mathsf{aff}}) \tag{2.87}$$

shows that the ternary relation Coll is positive-existentially definable from the $\mathcal{A}_{\mathsf{aff}}$-structure on $\mathbb{R}^2$.

We have the corresponding notions of homomorphisms for both structures:

2.88. **Definition.** A $\{\mathrm{Coll}\}$-homomorphism $T : \mathbb{R}^2 \to \mathbb{R}^2$ is called a **collineation**.

An $\mathcal{A}_{\mathsf{aff}}$-homomorphism $T : \mathbb{R}^2 \to \mathbb{R}^2$ is called an **affine transformation**.

2.89. **Corollary** (of Corollary 2.67/Exercise 2.79)**.** Every affine transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a collineation.

Proof. By the positive-existential definition (2.87). □

2.90. **Exercise.** Using the affine basis $(0,0),(1,0),(0,1)$ for $\mathbb{R}^2$, show that affine transformations of the plane are precisely all functions of the following form, for arbitrary $a,b,c,d,p,q \in \mathbb{R}$:

$$\begin{aligned} T : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ \begin{bmatrix} x \\ y \end{bmatrix} &\longmapsto \begin{bmatrix} ax+by+p \\ cx+dy+q \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} p \\ q \end{bmatrix}. \end{aligned}$$

2.91. **Exercise.**

(a) Give an example of a collineation $T : \mathbb{R}^2 \to \mathbb{R}^2$ which is not affine.

(b) Conclude that affine combinations are not positive-existentially definable from the collinearity relation (even using infinitary logic).

(c) (for set theorists) Show that there are strictly more collineations than affine transformations.

By an **invertible affine transformation**, respectively, **invertible collineation**, we mean an automorphism of $\mathbb{R}^2$ equipped with the respective structure. Note that since $\mathcal{A}_{\mathsf{aff}}$ consists only of function symbols, an invertible affine transformation is the same thing as a bijective one (by Proposition 2.77). The same happens to be true for collineations, even though Coll is a relation:

2.92. **Exercise.**

(a) Show that $A,B,C \in \mathbb{R}^2$ are *not* collinear iff they are distinct, and every point in $\mathbb{R}^2$ lies on a line through two distinct points each of which lies on one of the lines AB, BC, CA.

(b) Conclude that the inverse of every bijective collineation is a collineation.

Now comes the surprise:

2.93. **Theorem.** Every invertible collineation $T : \mathbb{R}^2 \to \mathbb{R}^2$ is affine.

In other words, keeping in mind the correspondence between automorphisms and (infinitary) definability (Exercise 2.81), this says that the purely qualitative notion of collinearity (the relation Coll) can in fact be used to define positions along a line (the operations AC_t)! Indeed, we will prove this theorem by giving such an explicit definition of AC_t from Coll. This will be done via a series of lemmas, each showing that progressively more "quantitative" notions can be defined from Coll.

2.94. **Lemma.** The quaternary relation

$$\mathrm{Para}(A,B,C,D) \;:\Longleftrightarrow\; A \neq B \text{ and } C \neq D \text{ and the lines } AB \text{ and } CD \text{ are parallel}$$

can be defined from Coll.

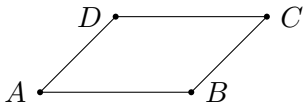
Proof. Two lines in the plane are parallel iff they don't intersect, or they are the same line; thus Para is defined by the {Coll}-formula

$$\neg(A=B) \wedge \neg(C=D) \wedge (\neg\exists E\,(\mathrm{Coll}(A,B,E) \wedge \mathrm{Coll}(C,D,E)) \vee (\mathrm{Coll}(A,B,C) \wedge \mathrm{Coll}(A,B,D))). \qquad \square$$

2.95. **Lemma.** The quaternary relation

$$\mathrm{Pgram}(A,B,C,D) \;:\Longleftrightarrow\; ABCD \text{ is a non-collinear parallelogram}$$

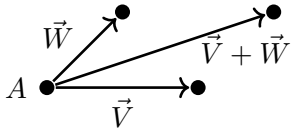
can be defined from Coll and Para, hence from just Coll.



Proof. We basically just need to say that opposite sides are parallel; for non-collinearity, it is enough to say that A,B,C are not collinear (which will in particular force all four vertices to be distinct, in agreement with the first two clauses in the definition of Para above):

$$\neg\mathrm{Coll}(A,B,C) \wedge \mathrm{Para}(A,B,C,D) \wedge \mathrm{Para}(A,D,B,C). \qquad \square$$

At this stage, we have already extracted a fair amount of quantitative information from the Coll relation. Indeed, note that parallelograms essentially allow us to express vector addition:



Informally speaking, all that is still needed in order to express arbitrary affine combinations is scalar multiplication of vectors. Scaling by an integer amount can be expressed via repeated addition; scaling by a rational a/b can then be expressed by saying that scaling one vector by a is equal to scaling another by b. Of course, there is the technical annoyance that we need to add *parallel* vectors here, while above we restricted to *non-collinear* parallelograms. This is easily worked around:

2.96. **Lemma.** The binary operation

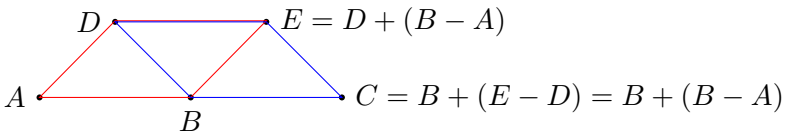
$$\begin{aligned} \mathrm{AC}_2 : (\mathbb{R}^2)^2 &\longrightarrow \mathbb{R}^2 \\ (A, B) &\longmapsto -A + 2B = B + (B - A) \end{aligned}$$

can be defined from Pgram, hence from Coll.

Proof. Recall that to say that the operation AC_2 is definable means that its ternary graph relation "$C = B + (B - A)$" is definable. Indeed, we have

$$C = B + (B - A) \iff ((A = B) \wedge (B = C)) \vee \exists D\, \exists E\, (\mathrm{Pgram}(A, B, E, D) \wedge \mathrm{Pgram}(B, C, E, D)).$$

When $A \neq B$, the existential expresses the following situation:



□

2.97. **Lemma.** For any rational $t \in \mathbb{Q}$, the operation $\mathrm{AC}_t : (\mathbb{R}^2)^2 \to \mathbb{R}^2$ is definable from Coll.

Proof. First, consider the case $t = n \in \mathbb{N}$. We use induction on n. For $n = 0$ or 1, we have

$$\mathrm{AC}_0(A, B) = A, \qquad \mathrm{AC}_1(A, B) = B.$$

For $n \geq 2$, by considering the picture

we are led to the calculation

$$\begin{aligned} \mathrm{AC}_n(A, B) &= (1 - n)A + nB \\ &= (2 - n)A + (n - 1)B + ((2 - n)A + (n - 1)B - (3 - n)A - (n - 2)B) \\ &= \mathrm{AC}_2(\mathrm{AC}_{n-2}(A, B), \mathrm{AC}_{n-1}(A, B)); \end{aligned}$$

since AC_2 is definable from Coll by the preceding lemma, while $\mathrm{AC}_{n-1}, \mathrm{AC}_{n-2}$ are definable by the IH, we get that AC_n is definable from Coll. This proves the case $t = n \in \mathbb{N}$.

For negative $t \in \mathbb{Z}$, we have

$$\mathrm{AC}_t(A, B) = (1 - t)A + tB = \mathrm{AC}_{1-t}(B, A);$$

since $t < 0$, $1 - t \geq 0$, so this reduces to the previous case. So we have proved all integer cases $t \in \mathbb{Z}$.

Finally, for a rational $t = \frac{p}{q}$ where $p, q \in \mathbb{Z}$, we have

$$\begin{aligned} \mathrm{AC}_t(A,B) = C &\iff (1-\tfrac{p}{q})A + \tfrac{p}{q}B = C \\ &\iff p(B-A) = q(C-A) \\ &\iff (1-p)A + pB = (1-q)A + qC \\ &\iff \mathrm{AC}_p(A,B) = \mathrm{AC}_q(A,C); \end{aligned}$$

since $\mathrm{AC}_p, \mathrm{AC}_q$ are definable from Coll by the previous cases, so is AC_t. □

In order to complete the proof, we just need to extend from rational t to arbitrary real $t \in \mathbb{R}$. This turns out to be *much* harder than any of the preceding lemmas. Intuitively, the idea is to "approximate" AC_t for an arbitrary real t by AC_r for rational $r \approx t$, and then take a limit as $r \to t$. We thus need to define "limit" from Coll. We can formulate this as follows:

2.98. **Lemma.** The ternary relation

$$\begin{aligned} \mathrm{Seg}(A,B,C) &:\iff C \text{ is on the line } \textit{segment } AB \\ &\iff C = \mathrm{AC}_t(A,B) \text{ for some } t \in [0,1] \end{aligned}$$

is definable from Coll. [Such affine combinations with $t \in [0,1]$ are called **convex combinations**.]

Once we know this lemma, we can approximate $t \in \mathbb{R}$ by rationals $r < t < s$ on either side of it, and then demand that $\mathrm{AC}_t(A,B)$ be between $\mathrm{AC}_r(A,B)$ and $\mathrm{AC}_s(A,B)$ for all such approximations; by taking r, s to be closer and closer to t, a "squeezing" argument then shows that $\mathrm{AC}_t(A,B)$ must be what it should be. Assuming Lemma 2.98 for now, we can thus complete the

Proof of Theorem 2.93. We claim that for any $t \in \mathbb{R}$,

$$(*) \qquad \mathrm{AC}_t(A,B) = C \iff \bigwedge_{\substack{r,s\in\mathbb{Q} \\ r<t<s}} \mathrm{Seg}(\mathrm{AC}_r(A,B), \mathrm{AC}_s(A,B), C).$$

This is enough, since $\mathrm{AC}_r, \mathrm{AC}_s$ are definable (in first-order logic) from Coll by Lemma 2.97, and Seg is definable by Lemma 2.98, whence AC_t is definable (in infinitary logic) from Coll.

To see $(*)$: if $\mathrm{AC}_t(A,B) = C$, then for all $r < t < s$, we have

$$\begin{aligned} C &= (1-t)A + tB \\ &= A + t(B-A) \\ &= A + (r + \tfrac{t-r}{s-r}(s-r))(B-A) \\ &= A + ((1-\tfrac{t-r}{s-r})r + \tfrac{t-r}{s-r}s)(B-A) \\ &= (1-\tfrac{t-r}{s-r})(A + r(B-A)) + \tfrac{t-r}{s-r}(A + s(B-A)) \\ &= (1-\tfrac{t-r}{s-r})((1-r)A + rB) + \tfrac{t-r}{s-r}((1-s)A + sB) \\ &= \mathrm{AC}_{\frac{t-r}{s-r}}(\mathrm{AC}_r(A,B), \mathrm{AC}_s(A,B)) \end{aligned}$$

where $\frac{t-r}{s-r} \in [0,1]$ because $r < t < s$, whence C is a convex combination of $\mathrm{AC}_r(A,B), \mathrm{AC}_s(A,B)$. Conversely, suppose that for all $r, s \in \mathbb{Q}$ with $r < t < s$, C is between $\mathrm{AC}_r(A,B), \mathrm{AC}_s(A,B)$. Since

$$\begin{aligned} \mathrm{AC}_s(A,B) - \mathrm{AC}_r(A,B) &= ((1-s)A + sB) - ((1-r)A + rB) \\ &= (s-r)(B-A), \end{aligned}$$

the distance between $\mathrm{AC}_r(A,B), \mathrm{AC}_s(A,B)$ is $s-r$ times the distance between A, B. Since C is between $\mathrm{AC}_r(A,B), \mathrm{AC}_s(A,B)$, as is $\mathrm{AC}_t(A,B)$ (as shown above), the distance between $C, \mathrm{AC}_t(A,B)$ is thus at most $s-r$ times the distance between A, B. Since we may choose r, s so that $s-r$ is arbitrarily small, the distance between $C, \mathrm{AC}_t(A,B)$ must thus be zero. □

So it remains to prove Lemma 2.98. We can make one more easy reduction:

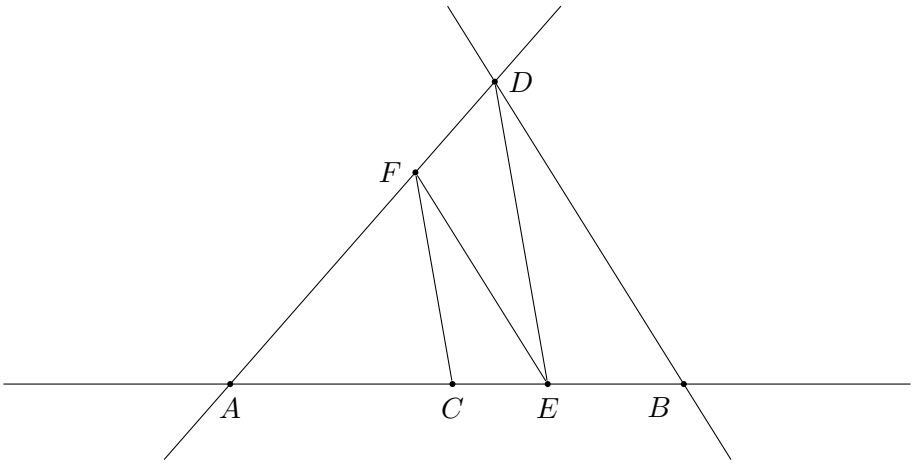
2.99. **Lemma.** The ternary relation

$$\mathrm{Ray}(A,B,C) \;:\Longleftrightarrow\; C = \mathrm{AC}_t(A,B) \text{ for some } t \in [0,\infty)$$

is definable from Coll.

Proof of Lemma 2.98. $\mathrm{Seg}(A,B,C) \iff \mathrm{Ray}(A,B,C) \wedge \mathrm{Ray}(B,A,C)$. □

Proof of Lemma 2.99. If $A = B$, then $C = \mathrm{AC}_t(A,B)$ again just means $A = B = C$. So we may restrict attention to the case $A \neq B$. Consider the following picture:



Here D is any point not on the line AB, while E is any point on the line AB, so that

$$E = (1-t)A + tB$$

for some $t \in \mathbb{R}$. The unique line through E parallel to BD intersects the line AD at a unique point F, since BD is not parallel to AD (since A, B, D are not collinear). We must have

$$F = (1-t)A + tD,$$

in order to ensure that the vector $F - E = t(D - B)$ is parallel to the line BD. Now similarly to before, the unique line through F parallel to DE intersects the line AB at a unique point C, since DE is not parallel to AB (since they intersect at E); and for the same reason as for F,

$$\begin{aligned} C &= (1-t)A + tE \\ &= (1-t)A + t((1-t)A + tB) \\ &= (1-t^2)A + t^2 B. \end{aligned}$$

The key thing to note is that the coefficient t^2 must be ≥ 0. In other words, what we have shown is that, starting from any point E on the *line* AB, and then constructing F and C uniquely as above, we end up with a point C on the *ray* AB; and this point C may be an arbitrary point $C = (1-s)A + sB$ on the ray AB, for any $s \geq 0$, since we may choose E above with $t := \sqrt{s}$. Putting everything together, we may define Ray from Coll and Para as follows:

$$\mathrm{Ray}(A,B,C) \iff (A = C) \vee \left(\neg(A = B) \wedge \exists D\, \exists E\, \exists F \begin{pmatrix} \neg\mathrm{Coll}(A,B,D) \wedge \mathrm{Coll}(A,B,E) \wedge \\ \mathrm{Para}(B,D,E,F) \wedge \mathrm{Coll}(A,D,F) \wedge \\ \mathrm{Para}(D,E,F,C) \wedge \mathrm{Coll}(A,C,B) \end{pmatrix} \right). \square$$

2.100. **Exercise.** Draw the above picture starting with E on the other side of A, to see that C still ends up on the right side of A.

3. First-order proofs

3.A. **Natural deduction.** We now define a natural deduction system for first-order logic. As in propositional logic, we will design the system so as to capture the informal proofs that mathematicians write in practice. Here is an example of an informal first-order proof:

3.1. **Example.** In every abelian group, for every x, y, there is a z such that $x + z = y$. $(\mathcal{T}_{\mathsf{abgrp}} \vdash \forall x \forall y \exists z (x+z=y))$

Proof. Let x, y be arbitrary; we must find z such that $x + z = y$. $(\mathcal{T}_{\mathsf{abgrp}} \vdash_{\{x,y\}} \exists z (x+z=y))$

We have

$$\left.\begin{aligned} x + ((-x) + y) &= (x + (-x)) + y && \text{by associativity} \\ &= 0 + y && \text{by inverse law} \\ &= y && \text{by identity law.} \end{aligned}\right\} (\mathcal{T}_{\mathsf{abgrp}} \vdash_{\{x,y\}} x+((-x)+y)=y)$$

Thus $z := (-x) + y$ works.

So indeed, for every x, y, there is a z such that $x + z = y$. □

Compared to propositional proofs, we see that, at each intermediate stage of the above proof, not only have we made some background assumptions (which in the above proof are always just the abelian group axioms, $\mathcal{T}_{\mathsf{abgrp}}$), but we may also have fixed some free variables (x, y above).

3.2. **Definition.** A **first-order $\mathcal{A}$-sequent** is an expression of the form

$$\mathcal{T} \vdash_X \phi,$$

read "$\mathcal{T}$ proves ϕ under X", where $\mathcal{T}$ is an $\mathcal{A}$-theory and ϕ is an $\mathcal{A}$-formula, both with free variables from X. Informally, X consists of some variables which have been "fixed" in the background, and $\mathcal{T}$ consists of the background assumptions which have been made so far.

In Example 3.1, we have labeled the sequents in three of the subproofs. We have not yet labeled any of the inference rules used to go between these sequents, because they all concern the new features of first-order formulas: quantifiers $\forall, \exists$ and equality $=$. (See Definition 3.5 below.)

Before introducing these rules, we need to address an annoying techniality. As in the above proof, we often need to *substitute* terms for variables in a first-order formula (e.g., $z \mapsto (-x) + y$ into $x + z = y$). This can cause clashes with bound variables:

3.3. **Example.** Consider the $\mathcal{A}_{\mathsf{ordfield}}$-formula

$$\phi := (x \leq y) \wedge \exists y\, (x + y = z).$$

Informally speaking, the two occurrences of y don't refer to the same thing: the second occurrence is bound by the $\exists y$, hence is "inaccessible from the outside". If we attempt to substitute $x \mapsto y$ into this formula, we get the "wrong" answer

$$\phi[x \mapsto y] = (y \leq y) \wedge \exists y\, (y + y = z).$$

Indeed, note for instance that while ϕ is true in $\mathbb{Z}$ under any assignment to x, z, the new formula is no longer true under $z \mapsto 1$! The "correct" substitution is to first change the bound variable in ϕ:

$$\phi' := (x \leq y) \wedge \exists w\, (x + w = z).$$

We may now safely substitute $x \mapsto y$, yielding

$$\phi'[x \mapsto y] = (y \leq y) \wedge \exists w\, (y + w = z).$$

Two formulas which differ only in bound variables such as ϕ, ϕ' above are called **α-equivalent**, denoted $\equiv_\alpha$; and it is a basic fact that we may always replace a formula by an α-equivalent copy to make a substitution safe. The proof of this fact, as well as the formal definition of $\equiv_\alpha$, is quite painful, and not really conceptually important at this point. We thus relegate it to the Appendix.

3.4. **Convention.** Henceforth, all formulas in sequents may be replaced with α-equivalent copies. In other words, sequents really consist of *α-equivalence classes* of formulas; however, we will continue to denote them as single formulas, as in $\{\phi, \psi\} \vdash_X \theta$, rather than $\{[\phi], [\psi]\} \vdash_X [\theta]$.

3.5. **Definition.** The **natural deduction system for first-order logic**, over the set of first-order sequents, has the following inference rule(schema)s:

- We include all the same rules as in propositional logic; this may be formalized via the following trick. For an alphabet $\mathcal{B}$ and propositional inference rule
$$\frac{\mathcal{T}_1 \vdash \phi_1 \quad \cdots \quad \mathcal{T}_n \vdash \phi_n}{\mathcal{T} \vdash \phi}$$
over $\mathcal{B}$-formulas, we include every **first-order instance** of it, obtained by performing the *same* formula substitution $\sigma : \mathcal{B} \to \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$ into all the formulas in it, resulting in
$$\frac{\mathcal{T}_1[\sigma] \vdash_X \phi_1[\sigma] \quad \cdots \quad \mathcal{T}_n[\sigma] \vdash_X \phi_n[\sigma]}{\mathcal{T}[\sigma] \vdash_X \phi[\sigma]}.$$

  3.6. **Example.**
$$(\vee\text{I1})\ \frac{\mathcal{T} \vdash_X \phi}{\mathcal{T} \vdash_X \phi \vee \psi} \quad \text{for } \mathcal{T} \subseteq \mathcal{L}^X_{\mathsf{form}}(\mathcal{A}),\ \phi, \psi \in \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$$
  is a first-order instance of the propositional ($\vee$I1) rule. *This is not completely obvious from the definition*: after all, the formulas ϕ, ψ, as well as the formulas in $\mathcal{T}$, may contain quantifiers. To see this in a systematic manner, for each $\theta \in \mathcal{T}$, let R_θ be an atomic propositional formula; then the above ($\vee$I1) is a first-order instance of the propositional rule
$$(\vee\text{I1})\ \frac{\{R_\theta \mid \theta \in \mathcal{T}\} \vdash P}{\{R_\theta \mid \theta \in \mathcal{T}\} \vdash P \vee Q}$$
  via the substitution $\sigma : \{R_\theta \mid \theta \in \mathcal{T}\} \cup \{P, Q\} \to \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$ mapping $R_\theta \mapsto \theta$, $P \mapsto \phi$, $Q \mapsto \psi$.

- For $=$, we have the following rules:
$$(=\text{I})\ \frac{}{\mathcal{T} \vdash_X t = t} \quad \text{for } t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A}),$$
$$(=\text{E})\ \frac{\mathcal{T} \vdash_X s = t \quad \mathcal{T} \vdash_X \phi[x \mapsto s]}{\mathcal{T} \vdash_X \phi[x \mapsto t]} \quad \text{for } s, t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A}),\ \phi \in \mathcal{L}^{X \cup \{x\}}_{\mathsf{form}}(\mathcal{A}) \text{ such that } \phi[x \mapsto s], \phi[x \mapsto t] \text{ are safe substitutions.}$$
  The (=I) rule says that everything equals itself, while the (=E) rule (sometimes called the **Leibniz rule**) says that things which are equal are interchangeable in every statement. We call ϕ here the **template formula** in which the two equal things can be swapped.

- Finally, for $\exists$, we have the following rules:
$$(\exists\text{I})\ \frac{\mathcal{T} \vdash_X \phi[x \mapsto t]}{\mathcal{T} \vdash_X \exists x\, \phi} \quad \text{for } \phi \in \mathcal{L}^{X \cup \{x\}}_{\mathsf{form}}(\mathcal{A}),\ t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A}) \text{ such that } \phi[x \mapsto t] \text{ is safe,}$$
$$(\exists\text{E})\ \frac{\mathcal{T} \vdash_X \exists x\, \phi \quad \mathcal{T} \cup \{\phi\} \vdash_{X \cup \{x\}} \psi}{\mathcal{T} \vdash_X \psi} \quad \text{for } \phi \in \mathcal{L}^{X \cup \{x\}}_{\mathsf{form}}(\mathcal{A}),\ \psi \in \mathcal{L}^X_{\mathsf{form}}(\mathcal{A}) \text{ such that } x \notin X.$$
  The ($\exists$I) rule says "to prove $\exists x\, \phi$, produce a witness t", while the ($\exists$E) rule says "to use $\exists x\, \phi$ to prove ψ, fix x such that ϕ, and prove ψ"; the restriction $x \notin X$ says that x is a *newly* fixed variable, about which the only thing we know is ϕ (since $\mathrm{FV}(\mathcal{T}) \subseteq X$).

The definitions of **derivable** and **admissible** inference rule are the same as in propositional logic: the former means there is a deduction of the given rule using only the basic rules above, while the latter means that deductions of the hypotheses of the given rule from *no* hypotheses may be transformed into a deduction of the conclusion from *no* hypotheses.

3.7. **Exercise.** Every first-order instance of every provable sequent in propositional logic is provable. ["Substitute σ into the deduction $\mathcal{D}$, yielding $\mathcal{D}[\sigma]$."]

For example, recall that for any propositional formula ϕ, it and its double negation $\neg\neg\phi$ provably imply each other (see Example 3.12 from propositional logic). Here was one of the deductions:

$$\dfrac{\dfrac{\dfrac{}{\{\phi,\neg\phi\}\vdash\phi}(\mathrm{A}) \quad \dfrac{}{\{\phi,\neg\phi\}\vdash\neg\phi}(\mathrm{A})}{\{\phi,\neg\phi\}\vdash\bot}(\neg\mathrm{E})}{\phi\vdash\neg\neg\phi}(\neg\mathrm{I})$$

We claim that the same holds for any *first-order* $\phi\in\mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$ (over the variables X). Again, this is perhaps not as obvious as it looks; we need to first take ϕ above to be an atomic formula P, and *then* perform the substitution $P\mapsto\phi$ for the desired first-order ϕ, in order to arrive at the deduction

$$\dfrac{\dfrac{\dfrac{}{\{\phi,\neg\phi\}\vdash_X\phi}(\mathrm{A}) \quad \dfrac{}{\{\phi,\neg\phi\}\vdash_X\neg\phi}(\mathrm{A})}{\{\phi,\neg\phi\}\vdash_X\bot}(\neg\mathrm{E})}{\phi\vdash_X\neg\neg\phi}(\neg\mathrm{I})$$

3.8. **Exercise.** Similarly, every first-order instance of a derivable propositional rule is derivable.

For example, the following rules are derivable:

$$(\mathrm{P})\ \dfrac{\mathcal{T}\cup\{\neg\phi\}\vdash_X\phi}{\mathcal{T}\vdash_X\phi} \qquad (\mathrm{LEM})\ \dfrac{}{\mathcal{T}\vdash_X\phi\vee\neg\phi} \qquad (\to\mathrm{I})\ \dfrac{\mathcal{T}\cup\{\phi\}\vdash_X\psi}{\mathcal{T}\vdash_X\phi\to\psi}$$

On the other hand, *it is not true that every first-order instance of an admissible propositional rule is automatically admissible*! Think about what this would mean: we would need to know that if all the hypotheses of the instance are provable, then so is the conclusion. *If* we knew that the proofs of the hypotheses of the instance were instances of proofs of the hypotheses of the original rule, *then* by admissibility of the original rule, its conclusion is provable, *hence* the conclusion of the instance rule is also provable. But a moment's thought reveals this to be an unreasonable "if":

3.9. **Example.** The propositional rule

$$\dfrac{\vdash P}{\vdash\bot}$$

is vacuously admissible: there is no proof of $\vdash P$ (by soundness, since P is not true under every truth assignment m). But the substitution $P\mapsto\top$ takes this to the instance

$$\dfrac{\vdash\top}{\vdash\bot}$$

which is no longer admissible, since there is a proof of $\vdash\top$ (by $(\top\mathrm{I})$, which is *not* an instance of any proof of $\vdash P$), but there is still no proof of $\vdash\bot$ (by soundness).

3.10. **Example.** The propositional rule

$$(\wedge\mathrm{I})\ \dfrac{\{P,Q\wedge R\}\vdash R \quad \{P,Q\wedge R\}\vdash S}{\{P,Q\wedge R\}\vdash R\wedge S}$$

has the first-order instance

$$(\wedge\mathrm{I})\ \dfrac{(0\le 1)\wedge(x\cdot y=0)\vdash_{\{x,y,z\}}x\cdot y=0 \quad (0\le 1)\wedge(x\cdot y=0)\vdash_{\{x,y,z\}}y\cdot z=1}{(0\le 1)\wedge(x\cdot y=0)\vdash_{\{x,y,z\}}(x\cdot y=0)\wedge(y\cdot z=1)}$$

via the substitution $\sigma:\{P,Q,R,S\}\to\mathcal{L}^{\{x,y,z\}}_{\mathsf{form}}(\mathcal{A}_{\mathsf{ordfield}})$ mapping $Q\mapsto 0\le 1$, $R\mapsto x\cdot y=0$ (which are the only possibilities for $\sigma(Q),\sigma(R)$, since $Q\wedge R$ needs to be mapped to a conjunction on the LHS), $S\mapsto y\cdot z=1$ (by considering the RHS of the second hypothesis), and $P\mapsto(0\le 1)\wedge(x\cdot y=0)$ (since P also needs to be mapped to some formula in the LHS theory).

On the other hand, the first-order rule

$$(\wedge\text{I})\ \frac{(0 \le 1) \wedge (x \cdot y = 0) \vdash_{\{x,y,z\}} 0 \le 1 \quad (0 \le 1) \wedge (x \cdot y = 0) \vdash_{\{x,y,z\}} y \cdot z = 1}{(0 \le 1) \wedge (x \cdot y = 0) \vdash_{\{x,y,z\}} (0 \le 1) \wedge (y \cdot z = 1)}$$

would *not* be a first-order instance of the above propositional rule, since R would need to be mapped to both $0 \le 1$ and $x \cdot y = 0$ (and these are not the same, even up to α-equivalence). Nonetheless, this rule *is* a first-order instance of a valid propositional ($\wedge$I), just not the one above; we just need to choose the propositional formulas in a more general fashion, e.g., by systematically assigning a different propositional symbol $P, Q, R, \ldots$ to each first-order formula, as explained in Definition 3.5:

$$(\wedge\text{I})\ \frac{P \vdash Q \quad P \vdash R}{P \vdash Q \wedge R}$$

which becomes the above rule under $P \mapsto (0 \le 1) \wedge (x \cdot y = 0)$, $Q \mapsto 0 \le 1$, and $R \mapsto y \cdot z = 1$.

Finally, neither of

$$(\wedge\text{I})\ \frac{(0 \le 1) \wedge (x \cdot y = 0) \vdash_{\{x,y\}} 0 \le 1 \quad (0 \le 1) \wedge (x \cdot y = 0) \vdash_{\{x,y\}} y \cdot z = 1}{(0 \le 1) \wedge (x \cdot y = 0) \vdash_{\{x,y\}} (0 \le 1) \wedge (y \cdot z = 1)},$$

$$(\wedge\text{I})\ \frac{(0 \le 1) \wedge (x \cdot y = 0) \vdash_{\{x,y\}} 0 \le 1 \quad (0 \le 1) \wedge (x \cdot y = 0) \vdash_{\{x,y,z\}} y \cdot z = 1}{(0 \le 1) \wedge (x \cdot y = 0) \vdash_{\{x,y,z\}} (0 \le 1) \wedge (y \cdot z = 1)}$$

is a valid first-order instance of any propositional rule. In the first rule, the second hypothesis as well as conclusion are not valid first-order sequents, since the fixed variable set $\{x, y\}$ does not include all the free variables appearing in the formulas on either side. The second rule does not have this problem; however, all the variable sets appearing in a first-order instance must be the same (and all the substitutions must be obtained via the same substitution σ).

We now turn to examples of the new first-order rules. The (=I) rule is self-explanatory. Here is an example of (=E); pay close attention to the role of the template formula:

3.11. **Example.** Here is part of the proof that $0 \le 1$ from the ordered field axioms (Example 2.31):

$$(=\text{E})\ \frac{\dfrac{\vdots}{\mathcal{T}_{\mathsf{ordfield}} \vdash 1 \cdot 1 = 1} \quad \dfrac{\vdots}{\mathcal{T}_{\mathsf{ordfield}} \vdash 0 \le 1 \cdot 1}}{\mathcal{T}_{\mathsf{ordfield}} \vdash 0 \le 1}$$

The template formula here is $\phi := (0 \le x)$, and we are plugging $x \mapsto 1 \cdot 1$ and $x \mapsto 1$ into it.

(The rest of the proof will have to wait until we have the ($\forall$E) rule, so that we can make use of the $\forall$ axioms in $\mathcal{T}_{\mathsf{ordfield}}$; see Exercise 3.30.)

3.12. **Example.** Here is another (more artificial) application of (=E):

$$(=\text{E})\ \frac{(\text{A})\ \dfrac{}{\{x = -y,\ x \le x\} \vdash_{\{x,y\}} x = -y} \quad (\text{A})\ \dfrac{}{\{x = -y,\ x \le x\} \vdash_{\{x,y\}} x \le x}}{\{x = -y,\ x \le x\} \vdash_{\{x,y\}} -y \le -y}$$

Here, the template formula *could* be $\phi := (x \le x)$, in which case we are substituting $x \mapsto x$ and $x \mapsto -y$ into it; but it is probably clearer to choose the template formula $\phi := (z \le z)$ instead, to emphasize that z is the "hole" in which we are replacing the equal terms $x = -y$. If we wanted to replace only *one* of the x's, we have no choice but to use a different variable in the template:

$$(=\text{E})\ \frac{(\text{A})\ \dfrac{}{\{x = -y,\ x \le x\} \vdash_{\{x,y\}} x = -y} \quad (\text{A})\ \dfrac{}{\{x = -y,\ x \le x\} \vdash_{\{x,y\}} x \le x}}{\{x = -y,\ x \le x\} \vdash_{\{x,y\}} -y \le x}$$

Here, the template formula is $\phi := (z \le x)$, with the substitutions $z \mapsto x$ and $z \mapsto -y$.

It is common to use (=E) indirectly, via one of the following familiar ways of reasoning about equality which are derived from (=E):

3.13. **Example.** The following **symmetry** rule for $=$

$$(\textsc{Sym})\ \frac{\mathcal{T} \vdash_X s = t}{\mathcal{T} \vdash_X t = s}$$

is derivable:

$$(=\text{E})\ \frac{\mathcal{T} \vdash_X s = t \qquad (=\text{I})\ \dfrac{}{\mathcal{T} \vdash_X s = s}}{\mathcal{T} \vdash_X t = s}$$

(Here we are applying (=E) to the template formula $x = s$ with substitutions $x \mapsto s, t$.)

3.14. **Exercise.** The following **transitivity** rule for $=$ is derivable:

$$(\textsc{Trans})\ \frac{\mathcal{T} \vdash_X r = s \quad \mathcal{T} \vdash_X s = t}{\mathcal{T} \vdash_X r = t}$$

More generally, for each $n \in \mathbb{N}$, the following rule is derivable:

$$(\textsc{Trans}_n)\ \frac{\mathcal{T} \vdash_X t_0 = t_1 \quad \mathcal{T} \vdash_X t_1 = t_2 \quad \cdots \quad \mathcal{T} \vdash_X t_{n-1} = t_n}{\mathcal{T} \vdash_X t_0 = t_n}$$

3.15. **Example.** The following **congruence** rule for $=$ is derivable:

$$(\textsc{Cong})\ \frac{\mathcal{T} \vdash_X s_1 = t_1 \quad \cdots \quad \mathcal{T} \vdash_X s_n = t_n}{\mathcal{T} \vdash_X f(s_1, \ldots, s_n) = f(t_1, \ldots, t_n)} \quad \text{for } f \in \mathcal{A}^n_{\mathsf{fun}}.$$

To prove this, we repeatedly apply (=E) (you should read this top-down):

$$(=\text{E})\ \frac{\mathcal{T} \vdash_X s_2 = t_2 \qquad (=\text{E})\ \dfrac{\mathcal{T} \vdash_X s_1 = t_1 \qquad (=\text{I})\ \dfrac{}{\mathcal{T} \vdash_X f(s_1, \ldots, s_n) = f(s_1, \ldots, s_n)}}{\mathcal{T} \vdash_X f(s_1, \ldots, s_n) = f(t_1, s_2, \ldots, s_n)}}{\dfrac{\mathcal{T} \vdash_X f(s_1, \ldots, s_n) = f(t_1, t_2, s_3, \ldots, s_n)}{\vdots}}$$

$$(=\text{E})\ \frac{\mathcal{T} \vdash_X s_n = t_n \qquad \vdots}{\mathcal{T} \vdash_X f(s_1, \ldots, s_n) = f(t_1, \ldots, t_n)}$$

3.16. **Exercise.** To which template ϕ and substitutions $x \mapsto s, t$ are we applying each (=E)?

3.17. **Example.** We can now formalize part of the informal proof from Example 3.1:

$$(\textsc{Trans}_3)\ \frac{\mathcal{D}_1 \qquad (\textsc{Cong})\ \dfrac{\mathcal{D}_2 = \dfrac{\vdots}{\mathcal{T}_{\mathsf{abgrp}} \vdash_{\{x,y\}} x + (-x) = 0} \qquad (=\text{I})\ \dfrac{}{\mathcal{T}_{\mathsf{abgrp}} \vdash_{\{x,y\}} y = y}}{\mathcal{T}_{\mathsf{abgrp}} \vdash_{\{x,y\}} (x + (-x)) + y = 0 + y} \qquad \mathcal{D}_3}{(\exists\text{I})\ \dfrac{\mathcal{T}_{\mathsf{abgrp}} \vdash_{\{x,y\}} x + ((-x) + y) = y}{\mathcal{T}_{\mathsf{abgrp}} \vdash_{\{x,y\}} \exists z\, (x + z = y)}}$$

The ($\exists$I) at the end is applied with the witness term t. To fill in the rest of the proof, we need the rules ($\forall$I) and ($\forall$E); see Example 3.27 below.

3.18. **Remark.** It is often helpful to think of $\exists$ as analogous to $\vee$: $\phi_1 \vee \phi_2$ means ϕ_i is true for some $i = 1, 2$, whereas $\exists x\, \phi$ means ϕ is true for some x in the underlying set (which may be infinite).

The ($\exists$I) rule is then analogous to ($\vee$I1) and ($\vee$I2) ("to prove $\exists x\, \phi$, prove ϕ for some x", namely the witness term t, versus "to prove $\phi_1 \vee \phi_2$, prove ϕ_i for some i").

Similarly, the ($\exists$E) rule is analogous to ($\vee$E) ("to use $\exists x\, \phi$ to prove ψ, prove ψ assuming ϕ for an arbitrary x" versus "to use $\phi_1 \vee \phi_2$ to prove ψ, prove ψ assuming ϕ_i for each $i = 1, 2$"). This is illustrated by the following example:

3.19. **Example.** In order to prove that in a field, an element with a reciprocal cannot be zero:

$$\dfrac{\dfrac{}{\mathcal{T}_{\mathsf{field}} \cup \{\exists y\,(x \cdot y = 1)\} \vdash_x \exists y\,(x \cdot y = 1)}\,(\mathrm{A}) \qquad \dfrac{\dfrac{\vdots}{\mathcal{T}_{\mathsf{field}} \cup \{\exists y\,(x \cdot y = 1),\, x \cdot y = 1,\, x = 0\} \vdash_{\{x,y\}} \bot}}{\mathcal{T}_{\mathsf{field}} \cup \{\exists y\,(x \cdot y = 1),\, x \cdot y = 1\} \vdash_{\{x,y\}} \neg(x = 0)}\,(\neg\mathrm{I})}{\mathcal{T}_{\mathsf{field}} \cup \{\exists y\,(x \cdot y = 1)\} \vdash_x \neg(x = 0)}\,(\exists\mathrm{E})$$

Note how this application of (∃E) satisfies the condition $y \notin \{x\}$ from Definition 3.5, guaranteeing that y is a newly fixed variable about which nothing else is known.

3.20. **Exercise.** Finish the proof. [Use Exercise 3.29.]

3.21. **Example.** Here is an *invalid* application of (∃E), showing what could go wrong when we forget to check the condition that the variable is new:

$$\dfrac{\dfrac{}{\exists x\,(x = 1) \vdash_x \exists x\,(x = 1)}\,(\mathrm{A}) \qquad \dfrac{}{\{\exists x\,(x = 1),\, x = 1\} \vdash_x x = 1}\,(\mathrm{A})}{\exists x\,(x = 1) \vdash_x x = 1}\,(\exists\mathrm{E})$$

Indeed, the conclusion says that under the assumption $\exists x\,(x = 1)$ (true in any structure with a constant 1), we should have $x = 1$ for an arbitrary x (clearly not true in general).

The correct way to apply (∃E) in this deduction is to first change $\exists x\,(x = 1)$ to α-equivalent formula using a new bound variable:

$$\dfrac{\dfrac{}{\{\exists x\,(x = 1)\} \vdash_x \exists x\,(x = 1) \equiv_\alpha \exists y\,(y = 1)}\,(\mathrm{A}) \qquad \dfrac{???}{\{\exists x\,(x = 1),\, y = 1\} \vdash_{\{x,y\}} x = 1}}{\{\exists x\,(x = 1)\} \vdash_x x = 1}\,(\exists\mathrm{E})$$

We can no longer complete the proof, as expected.

3.22. **Example.** Let us show that for any $\phi, \psi \in \mathcal{L}_{\mathsf{form}}^{X \cup \{x\}}(\mathcal{A})$, we have the provable equivalence

$$\vdash_X (\exists x\,(\phi \vee \psi)) \leftrightarrow (\exists x\,\phi) \vee (\exists x\,\psi).$$

After applying (∧I) and (→I) as usual, this amounts to proving

$$\{\exists x\,(\phi \vee \psi)\} \vdash_X (\exists x\,\phi) \vee (\exists x\,\psi), \qquad\qquad \{(\exists x\,\phi) \vee (\exists x\,\psi)\} \vdash_X \exists x\,(\phi \vee \psi).$$

To prove the first sequent, we would naturally want to

> "Fix x such that $\phi \vee \psi$ holds, and then split into the cases where ϕ or ψ holds, in each case proving the respective clause on the RHS."

The only snag is that x could already have been fixed, i.e., maybe $x \in X$. To circumvent this, pick some $y \notin X$ which is also not bound in any of the above formulas, so that $(\phi \vee \psi)[x \mapsto y]$ is a safe substitution (see Appendix) and so $\exists x\,(\phi \vee \psi) \sim_\alpha \exists y\,(\phi \vee \psi)[x \mapsto y]$. We can then proceed to formalize the above proof sketch:

$$\dfrac{\dfrac{}{\exists x\,(\phi \vee \psi) \vdash_X \overbrace{\exists y\,(\phi \vee \psi)[x \mapsto y]}^{\exists x\,(\phi \vee \psi) \equiv_\alpha}}\,(\mathrm{A}) \qquad \dfrac{\dfrac{}{\mathcal{T} \vdash_{X \cup \{y\}} (\phi \vee \psi)[x \mapsto y]}\,(\mathrm{A}) \qquad \dfrac{\dfrac{\dfrac{}{\mathcal{T} \cup \{\phi[x \mapsto y]\} \vdash_{X \cup \{y\}} \phi[x \mapsto y]}\,(\mathrm{A})}{\mathcal{T} \cup \{\phi[x \mapsto y]\} \vdash_{X \cup \{y\}} \exists x\,\phi}\,(\exists\mathrm{I})}{\mathcal{T} \cup \{\phi[x \mapsto y]\} \vdash_{X \cup \{y\}} (\exists x\,\phi) \vee (\exists x\,\psi)}\,(\vee\mathrm{I1}) \quad \vdots}{\mathcal{T} := \{\exists x\,(\phi \vee \psi),\, (\phi \vee \psi)[x \mapsto y]\} \vdash_{X \cup \{y\}} (\exists x\,\phi) \vee (\exists x\,\psi)}\,(\vee\mathrm{E})}{\exists x\,(\phi \vee \psi) \vdash_X (\exists x\,\phi) \vee (\exists x\,\psi)}\,(\exists\mathrm{E})$$

where the last missing sub-deduction on the right is similar to the one to its left, and where in applying (∃I) we're using that $\phi[x \mapsto y]$ is safe since $(\phi \vee \psi)[x \mapsto y]$ is.

3.23. **Exercise.** Give the deduction of the converse $\{(\exists x\,\phi) \vee (\exists x\,\psi)\} \vdash_X \exists x\,(\phi \vee \psi)$.

In order to give more interesting examples of first-order deductions, we now need

3.24. **Proposition** (weakening)**.** The following rule is admissible, for $\mathcal{T} \subseteq \mathcal{T}' \subseteq \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$:

$$(\mathrm{W})\ \frac{\mathcal{T} \vdash_X \phi}{\mathcal{T}' \vdash_X \phi}$$

(Note that for now, the variable set has to remain the same; see also Corollary 3.33 below.)

Proof. By induction on the deduction of $\mathcal{T} \vdash_X \phi$, similarly to the propositional case (see Proposition 3.17 from propositional logic); the point is that in all of the new first-order inference rules in Definition 3.5, we may also freely introduce extra assumptions into the theory. □

3.25. **Example.** It follows that *every first-order instance of an admissible propositional rule which was derived using weakening is now admissible*. This is because we may perform the formula substitution into the deduction of the propositional rule, as in Exercise 3.8 for derivable rules, and then replace all the resulting first-order instances of propositional (W) using the first-order (W) above. For example, we get that the following first-order rules are admissible:

$$(\to\mathrm{E})\ \frac{\mathcal{T} \vdash_X \phi \to \psi \quad \mathcal{T} \vdash_X \phi}{\mathcal{T} \vdash_X \psi} \qquad (\textsc{Cut})\ \frac{\mathcal{T} \vdash_X \phi \quad \mathcal{T} \cup \{\phi\} \vdash_X \psi}{\mathcal{T} \vdash_X \psi}$$

3.26. **Example.** Recalling that $\forall x$ is an abbreviation for $\neg\exists x\, \neg$, we have the admissible rules for $\forall$ ("to prove $\forall x\, \phi$, let x be arbitrary and prove ϕ", and "from $\forall x\, \phi$, deduce ϕ for a particular x"):

$$(\forall\mathrm{I})\ \frac{\mathcal{T} \vdash_{X\cup\{x\}} \phi}{\mathcal{T} \vdash_X \forall x\, \phi}\quad \begin{array}{l}\text{for } \phi \in \mathcal{L}^{X\cup\{x\}}_{\mathsf{form}}(\mathcal{A})\\ \text{such that } x \notin X,\end{array}$$

$$(\forall\mathrm{E})\ \frac{\mathcal{T} \vdash_X \forall x\, \phi}{\mathcal{T} \vdash_X \phi[x \mapsto t]}\quad \begin{array}{l}\text{for } \phi \in \mathcal{L}^{X\cup\{x\}}_{\mathsf{form}}(\mathcal{A}),\ t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})\\ \text{such that } \phi[x \mapsto t] \text{ is safe.}\end{array}$$

Note the similarity to the rules for $\exists$ but with intro and elim swapped. As in Remark 3.18, it is helpful to think of these rules as analogous to those for $\wedge$: $\forall$ is like a conjunction indexed by the underlying set. Often more helpful, however, is a *different* analogy: $\forall$ is analogous to $\to$ ("if ..., then ...", versus "if we have some x, then ..."). This latter analogy is especially evident in the inference rules: both sets of rules are derived/admissible; the ($\to$I) and ($\forall$I) rules are clear parallels ("assume ...", versus "assume x is arbitrary"); and we can also view ($\to$E) and ($\forall$E) as parallels ("prove ..., and deduce ...", versus "produce an x, and deduce ... for it").

The ($\forall$I) rule is derivable using (W), hence admissible:

$$\dfrac{(\mathrm{A})\ \dfrac{}{\mathcal{T} \cup \{\exists x\, \neg\phi\} \vdash_X \exists x\, \neg\phi} \qquad (\mathrm{W})\ \dfrac{\mathcal{T} \vdash_{X\cup\{x\}} \phi}{\mathcal{T}' \vdash_{X\cup\{x\}} \phi} \quad (\mathrm{A})\ \dfrac{}{\mathcal{T}' \vdash_{X\cup\{x\}} \neg\phi}}{(\exists\mathrm{E})\ \dfrac{}{(\neg\mathrm{E})\ \dfrac{}{\mathcal{T}' := \mathcal{T} \cup \{\exists x\, \neg\phi, \neg\phi\} \vdash_{X\cup\{x\}} \bot}}}$$
$$(\neg\mathrm{I})\ \frac{\mathcal{T} \cup \{\exists x\, \neg\phi\} \vdash_X \bot}{\mathcal{T} \vdash_X \neg\exists x\, \neg\phi}$$

Note that the condition on x in ($\forall$I) implies the required condition in the application of ($\exists$E).

The ($\forall$E) rule is also derivable using (W), hence admissible:

$$\dfrac{(\mathrm{A})\ \dfrac{}{(\exists\mathrm{I})\ \dfrac{}{\mathcal{T} \cup \{\neg\phi[x \mapsto t]\} \vdash_X \neg\phi[x \mapsto t]}}}{(\neg\mathrm{E})\ \dfrac{}{\mathcal{T} \cup \{\neg\phi[x \mapsto t]\} \vdash_X \exists x\, \neg\phi}} \qquad (\mathrm{W})\ \frac{\mathcal{T} \vdash_X \neg\exists x\, \neg\phi}{\mathcal{T} \cup \{\neg\phi[x \mapsto t]\} \vdash_X \neg\exists x\, \neg\phi}$$
$$(\mathrm{C})\ \frac{\mathcal{T} \cup \{\neg\phi[x \mapsto t]\} \vdash_X \bot}{\mathcal{T} \vdash_X \phi[x \mapsto t]}$$

Again, the conditions in ($\forall$E) imply the required conditions in ($\exists$I).

3.27. **Example.** To finish Example 3.17 (formalizing Example 3.1):

$$\dfrac{\dfrac{\dfrac{\vdots}{\mathcal{T}_{\mathsf{abgrp}} \vdash_{\{x,y\}} \exists z\,(x+z=y)}(\forall\mathrm{I})}{\mathcal{T}_{\mathsf{abgrp}} \vdash_{\{x\}} \forall y\,\exists z\,(x+z=y)}(\forall\mathrm{I})}{\mathcal{T}_{\mathsf{abgrp}} \vdash_{\varnothing} \forall x\,\forall y\,\exists z\,(x+z=y)}, \qquad \mathcal{D}_2 = \dfrac{\dfrac{}{\mathcal{T}_{\mathsf{abgrp}} \vdash_{\{x,y\}} \forall x\,(x+(-x)=0)}(\mathrm{A})}{\mathcal{T}_{\mathsf{abgrp}} \vdash_{\{x,y\}} x+(-x)=0}(\forall\mathrm{E}).$$

3.28. **Exercise.** Fill in $\mathcal{D}_1, \mathcal{D}_3$ in Example 3.17.

3.29. **Exercise.** Give deductions of

$$\mathcal{T}_{\mathsf{commring}} \vdash_x x \cdot 0 = 0, \qquad \mathcal{T}_{\mathsf{commring}} \vdash \forall x\,\forall y\,(x \cdot -y = -(x \cdot y)).$$

[Formalize Exercise 2.32.] Using this, give a deduction of

$$\mathcal{T}_{\mathsf{commring}} \vdash \forall x\,(-(-x) = x).$$

[Informal proof: $-(-x) = 0 + (-(-x)) = x + (-x) + (-(-x)) = x + 0 = x$.]

3.30. **Exercise.** Formalize the following statement into a sequent, and give a deduction of it:

"In an ordered field, the square of every element is ≥ 0."

[Informal proof: since $\leq$ is a total order, every x is either ≤ 0 or ≥ 0. If $x \geq 0$, then since multiplication by nonnegative elements is order-preserving (by one of the ordered field axioms), we get $x \cdot x \geq 0 \cdot x = 0$. If $x \leq 0$, then adding $-x$ to both sides yields $0 \leq -x$, whence by the previous case, $0 \leq (-x)^2 = x^2$.]

Using this (and the identity axiom for $\cdot$), finish Example 3.11.

3.B. **Rules for variables.** We now discuss some additional admissible rules relating to free variables. Like most such discussions (see Appendix on variable substitution), the details are a bit tedious.

3.31. **Proposition** (substitution)**.** The following rule is admissible, for any variable substitution $\sigma : X \to \mathcal{L}^Y_{\mathsf{term}}(\mathcal{A})$ such that the substitutions $\mathcal{T}[\sigma], \phi[\sigma]$ are safe:

$$(\mathrm{S})\ \frac{\mathcal{T} \vdash_X \phi}{\mathcal{T}[\sigma] \vdash_Y \phi[\sigma]}$$

3.32. **Example.** From Example 3.17, we have $\mathcal{T}_{\mathsf{abgrp}} \vdash_{\{x,y\}} \exists z\,(x+z=y)$. By (S) with $x \mapsto y$, we get $\mathcal{T}_{\mathsf{abgrp}} \vdash_{\{y\}} \exists z\,(y+z=y)$. (Note that since $\mathcal{T}_{\mathsf{abgrp}}$ consists of sentences, $\mathcal{T}_{\mathsf{abgrp}}[\sigma] = \mathcal{T}_{\mathsf{abgrp}}$.)

On the other hand, if we had admissibility of (S) with $x \mapsto z$ (which is not safe for substitution into $\exists z\,(x+z=y)$), we would get $\mathcal{T}_{\mathsf{abgrp}} \vdash_{\{y\}} \exists z\,(z+z=y)$, which would violate soundness (Proposition 3.37) since the abelian group $\mathbb{Z}$ with the variable assignment $y \mapsto 1$ fails to satisfy this formula.

As a special case of (S), we can take $X \subseteq Y$, and take σ to be the identity function $X \to X \subseteq Y$, substitution of which is always safe, yielding

3.33. **Corollary** (variable weakening)**.** The following rule is admissible, for $X \subseteq Y \subseteq \mathcal{V}$:

$$(\mathrm{S})\ \frac{\mathcal{T} \vdash_X \phi}{\mathcal{T} \vdash_Y \phi}$$

Intuitively, this says that if we can prove ϕ after fixing some variables, the same proof should still apply after fixing some more extraneous variables. This is analogous to the "ordinary" weakening rule (W) (Proposition 3.24), which says that we can add some extraneous assumptions.

Conversely, we have:

3.34. **Proposition** (syntactic compactness)**.** If $\mathcal{T} \vdash_X \phi$, then there are finite $\mathcal{T}' \subseteq \mathcal{T}$ and $X' \subseteq X$ such that (X' contains all free variables occurring in $\mathcal{T}', \phi$, and) $\mathcal{T}' \vdash_{X'} \phi$.

Proof. First, we prove that keeping X fixed, we may shrink $\mathcal{T}$ down to a finite $\mathcal{T}' \subseteq \mathcal{T}$. This is exactly the same as in propositional logic (Proposition 3.26 in notes), by induction on the deduction of $\mathcal{T} \vdash_X \phi$, using that each step of the proof only uses at most one formula in $\mathcal{T}$.

It thus suffices to assume that $\mathcal{T}$ is already finite to begin with, and prove that we may shrink X down to a finite $X' \subseteq X$ containing all the free variables in $\mathcal{T}$. We proceed by induction on the deduction of $\mathcal{T} \vdash_X \phi$.

- If the deduction ends with (A), then $X' := \mathrm{FV}(\mathcal{T}) \cup \mathrm{FV}(\phi)$ works.
- If the deduction ends with, say,
$$(\vee\mathrm{E})\ \frac{\mathcal{T} \vdash_X \phi \vee \psi \quad \mathcal{T} \cup \{\phi\} \vdash_X \theta \quad \mathcal{T} \cup \{\psi\} \vdash_X \theta}{\mathcal{T} \vdash_X \theta},$$
then by the IH, there are finite $X_1, X_2, X_3 \subseteq X$ such that
$$\mathcal{T} \vdash_{X_1} \phi \vee \psi, \qquad \mathcal{T} \cup \{\phi\} \vdash_{X_2} \theta, \qquad \mathcal{T} \cup \{\psi\} \vdash_{X_3} \theta.$$
Let $X' := X_1 \cup X_2 \cup X_3$. Then by variable weakening, we may replace X_1, X_2, X_3 above with X', whence by ($\vee$E), $\mathcal{T} \vdash_{X'} \theta$.
- The rest of the first-order instances of propositional inference rules are similarly handled.
- If the deduction ends with $(=\mathrm{I})\ \dfrac{}{\mathcal{T} \vdash_X t = t}$, then $X' := \mathrm{FV}(\mathcal{T}) \cup \mathrm{FV}(t)$ works.
- If the deduction ends with
$$(=\mathrm{E})\ \frac{\mathcal{T} \vdash_X s = t \quad \mathcal{T} \vdash_X \phi[x \mapsto s]}{\mathcal{T} \vdash_X \phi[x \mapsto t]}$$
where $s, t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$ with $\phi[x \mapsto s], \phi[x \mapsto t]$ safe, then similarly to the ($\vee$E) case above, we may find a finite $X' \subseteq X$ such that
$$\mathcal{T} \vdash_{X'} s = t, \qquad \mathcal{T} \vdash_{X'} \phi[x \mapsto s];$$
in particular, for the first sequent to make sense, we must have $s, t \in \mathcal{L}^{X'}_{\mathsf{term}}(\mathcal{A})$, so that we may apply (=E) to get $\mathcal{T} \vdash_{X'} \phi[x \mapsto t]$.
- The ($\exists$I) case is similar (except that we should explicitly include all free variables in the witness term into X', to make sure we are still allowed to apply ($\exists$I)).
- Finally, if the deduction ends with
$$(\exists\mathrm{E})\ \frac{\mathcal{T} \vdash_X \exists x\, \phi \quad \mathcal{T} \cup \{\phi\} \vdash_{X \cup \{x\}} \psi}{\mathcal{T} \vdash_X \psi}$$
with $x \notin X$, then by the IH, there are finite $X_1 \subseteq X$ and $X_2 \subseteq X \cup \{y\}$ such that
$$\mathcal{T} \vdash_{X_1} \exists x\, \phi, \qquad \mathcal{T} \cup \{\phi\} \vdash_{X_2} \psi.$$
Let $X' := X_1 \cup (X_2 \cap X)$ (cf. the proof of syntactic compactness, Proposition 3.21, from propositional logic). Then $X_2 \subseteq X' \cup \{y\}$ (because $X_2 \subseteq X \cup \{x\}$), so by variable weakening,
$$\mathcal{T} \vdash_{X'} \exists x\, \phi, \qquad \mathcal{T} \cup \{\phi\} \vdash_{X' \cup \{x\}} \psi.$$
Moreover, $x \notin X'$ since $x \notin X \supseteq X'$, so we may apply ($\exists$E) to deduce $\mathcal{T} \vdash_{X'} \psi$. □

In the rest of this subsection, we give the proof of Proposition 3.31. This proof is rather technical, but the basic idea is straightforward enough: we should be able to simply perform the variable substitution σ throughout the entire deduction $\mathcal{D}$ of $\mathcal{T} \vdash_X \phi$. This is analogous to how one performs a *formula* substitution into a *propositional* deduction, as in Exercise 3.7.

The added difficulties are because, when substituting into a *first-order* sequent, one expects to encounter issues with variable capture:

- Even though in Proposition 3.31, we assumed that the substitution of σ into the conclusion $\mathcal{T} \vdash_X \phi$ of $\mathcal{D}$ is safe, this does *not* ensure that the substitution into every formula in $\mathcal{D}$ is safe, since $\mathcal{D}$ could contain complicated intermediate formulas that don't appear anywhere in its conclusion (for example, in the hypotheses of ($\vee$E) or ($\exists$E)).
- A more serious issue is that even if the substitution of σ into every *formula* in $\mathcal{D}$ is safe, we could still end up with an invalid deduction, *because the resulting deduction may violate the condition $x \notin X$ in* ($\exists$E). Indeed, this can happen even in the special case where σ is the identity, i.e., we are performing variable weakening. For example, in trying to weaken

$$(\exists\text{E})\ \frac{\begin{array}{cc}\vdots & \vdots\\ \cdots \vdash_{\{x\}} \exists y\,(x\cdot y = 1) & \cdots \cup \{x\cdot y = 1\} \vdash_{\{x,y\}} \neg(x=0)\end{array}}{\cdots \vdash_{\{x\}} \neg(x=0)}$$

to the bigger set of variables $\{x, y\} \supseteq \{x\}$, we obtain the invalid

$$(\exists\text{E})\ \frac{\begin{array}{cc}\vdots & \vdots\\ \cdots \vdash_{\{x,y\}} \exists y\,(x\cdot y = 1) & \cdots \cup \{x\cdot y = 1\} \vdash_{\{x,y\}} \neg(x=0)\end{array}}{\cdots \vdash_{\{x,y\}} \neg(x=0)}.$$

To handle this second type of issue, we say that **the substitution of $\sigma : X \to \mathcal{L}^Y_{\mathsf{term}}(\mathcal{A})$ into $\mathcal{D}$ is safe** if in no application of ($\exists$E) in $\mathcal{D}$ is the added variable x in Y.

3.35. **Lemma.** If $\mathcal{T} \vdash_X \phi$ via a deduction into which substitution of $\sigma : X \to \mathcal{L}^Y_{\mathsf{term}}(\mathcal{A})$ is safe, then $\mathcal{T}[\sigma] \vdash_Y \phi[\sigma]$.

Proof. By induction on the deduction $\mathcal{D}$ of $\mathcal{T} \vdash_X \phi$.

- The first-order instances of propositional rules are all straightforward: just substitute σ safely into all formulas in sight, possibly after replacing them with α-equivalent copies.
- If $\mathcal{D} = (=\text{I})\ \dfrac{}{\mathcal{T} \vdash_X t = t}$, then $\mathcal{T}[\sigma] \vdash_Y (t=t)[\sigma] = (t[\sigma] = t[\sigma])$ again by (=I).
- Suppose $\mathcal{D}$ ends with
$$(=\text{E})\ \frac{\mathcal{T} \vdash_X s = t \quad \mathcal{T} \vdash_X \phi[x \mapsto s]}{\mathcal{T} \vdash_X \phi[x \mapsto t]}$$
where both substitutions are safe. After substituting σ into everything in sight, we get
$$(*)\qquad (=\text{E})\ \frac{\mathcal{T}[\sigma] \vdash_Y s[\sigma] = t[\sigma] \quad \mathcal{T}[\sigma] \vdash_Y \phi[x \mapsto s][\sigma]}{\mathcal{T}[\sigma] \vdash_Y \phi[x \mapsto t][\sigma]}.$$
In order for this to be a valid application of (=E), we need for $\phi[x \mapsto s][\sigma]$ and $\phi[x \mapsto t][\sigma]$ to be substitutions of $s[\sigma], t[\sigma]$ into some common template formula. Because of the way a safe double substitution behaves (Exercise 3.4 from the Appendix on substitution),
$$\phi[x \mapsto s][\sigma] = \phi[(x \mapsto s)[\sigma]] = \phi[\sigma\langle x \mapsto s[\sigma]\rangle] = \phi[\sigma\langle x \mapsto y\rangle][y \mapsto s[\sigma]]$$
where y is a *new* variable not in Y, hence not in any term in the image of σ, in order to ensure $\sigma\langle x \mapsto y\rangle[y \mapsto s[\sigma]] = \sigma\langle x \mapsto s[\sigma]\rangle$. Moreover, this last substitution is also safe, provided we first replace ϕ with an α-equivalent formula none of whose bound variables appear free in Y (hence in $s[\sigma]$). Similarly,
$$\phi[x \mapsto t][\sigma] = \phi[(x \mapsto t)[\sigma]] = \phi[\sigma\langle x \mapsto t[\sigma]\rangle] = \phi[\sigma\langle x \mapsto y\rangle][y \mapsto t[\sigma]].$$
Thus $(*)$ is indeed a valid application of (=E), to the template formula $\phi[\sigma\langle x \mapsto y\rangle]$.

- Suppose $\mathcal{D}$ ends with

$$(\exists\mathrm{I})\ \frac{\mathcal{T} \vdash_X \phi[x \mapsto t]}{\mathcal{T} \vdash_X \exists x\, \phi}$$

where $\phi[x \mapsto t]$ is safe. After substituting σ into everything in sight, we get

$$(\dagger) \qquad (\exists\mathrm{I})\ \frac{\mathcal{T} \vdash_Y \phi[x \mapsto t][\sigma]}{\mathcal{T}[\sigma] \vdash_Y \exists x\, \phi[\sigma\langle x \mapsto x\rangle]}$$

Similarly to the (=E) case above, we have

$$\phi[x \mapsto t][\sigma] = \phi[(x \mapsto t)[\sigma]] = \phi[\sigma\langle x \mapsto t[\sigma]\rangle] = \phi[\sigma\langle x \mapsto y\rangle][y \mapsto t[\sigma]]$$

where y is chosen as above. Now in order for (†) to be a valid application of (∃I), we note that the formula in its conclusion is

$$\exists x\, \phi[\sigma\langle x \mapsto x\rangle] \sim_\alpha \exists y\, \phi[\sigma\langle x \mapsto x\rangle][x \mapsto y] = \exists y\, \phi[\sigma\langle x \mapsto y\rangle],$$

using that the substitutions $\sigma\langle x \mapsto x\rangle[x \mapsto y]$ and $\sigma\langle x \mapsto y\rangle$ agree on all free variables in ϕ: they clearly agree on x; while for $z \in \mathrm{FV}(\phi) \setminus \{x\}$, we have $\sigma(z)[x \mapsto y] = \sigma(z)$ since $x \notin \mathrm{FV}(\sigma(z))$ by the original assumption (in the statement of (S)) that $(\exists x\, \phi)[\sigma]$ is safe.

- Finally, suppose $\mathcal{D}$ ends with

$$(\exists\mathrm{E})\ \frac{\mathcal{T} \vdash_X \exists x\, \phi \quad \mathcal{T} \cup \{\phi\} \vdash_{X \cup \{x\}} \psi}{\mathcal{T} \vdash_X \psi}$$

where $x \notin X$. This means $\sigma\langle x \mapsto x\rangle$ is simply σ extended by the identity function (without erasing any previous $\sigma(x)$). Thus, substituting σ into everything yields

$$(\exists\mathrm{E})\ \frac{\mathcal{T}[\sigma] \vdash_Y \exists x\, \phi[\sigma] \quad \mathcal{T}[\sigma] \cup \{\phi[\sigma]\} \vdash_{Y \cup \{x\}} \psi[\sigma]}{\mathcal{T}[\sigma] \vdash_Y \psi[\sigma]}$$

which is still a valid application of (∃E) since $x \notin Y$ by our assumption that substitution of σ into $\mathcal{D}$ is safe. □

In order to finish proving Proposition 3.31, it thus remains to show that any deduction of $\mathcal{T} \vdash_X \phi$ can be turned into one into which substitution of σ is safe. The idea here is to think of the rule

$$(\exists\mathrm{E})\ \frac{\begin{array}{c}\vdots\\ \mathcal{T} \vdash_X \exists x\, \phi\end{array} \quad \begin{array}{c}\vdots\\ \mathcal{T} \cup \{\phi\} \vdash_{X \cup \{x\}} \psi\end{array}}{\mathcal{T} \vdash_X \psi}$$

as "binding" the free variable x in the second sub-deduction, much as a quantifier $\exists x\, \phi$ binds the free x in the subformula ϕ. When a substitution into this rule breaks the condition $x \notin X$, we should think of the (∃E) as "capturing" the free variable x. The solution to variable capture is the familiar one: we need to replace the original deduction with an "α-equivalent" copy, where the free variable has been replaced with a new variable via a safe substitution. Since this is essentially similar to the arguments for α-equivalence of formulas in the Appendix, the details are left to you:

3.36. **Exercise.**

(a) Prove by induction that if $\mathcal{T} \vdash_X \phi$, then for any infinite set Y disjoint from X, there is a deduction of $\mathcal{T} \vdash_X \phi$ whose new variables introduced by applications of (∃E) all come from Y. [Imitate the proof of Proposition 4.11 from the Appendix.]

(b) Conclude that for any $\sigma : X \to \mathcal{L}^Y_{\mathsf{term}}(\mathcal{A})$, if $\mathcal{T} \vdash_X \phi$, then there is a deduction of it into which substitution of σ is safe. Thereby conclude Proposition 3.31.

3.C. **Soundness and completeness.** Recalling Remark 2.40, for a theory $\mathcal{T}$ and formula ϕ both with free variables from X, we of course call ϕ a **semantic consequence** of $\mathcal{T}$ if

$$\mathcal{T} \models_X \phi \;:\iff\; (\mathcal{M}, \alpha) \in \mathrm{Mod}_X(\mathcal{T}) \iff \forall \mathcal{M}, \forall \alpha : X \to M,\ \mathcal{M} \models_\alpha \phi.$$

3.37. **Proposition** (soundness)**.** If $\mathcal{T} \vdash_X \phi$, then $\mathcal{T} \models_X \phi$.

Proof. We assume that there is a deduction $\mathcal{D}$ of $\mathcal{T} \vdash_X \phi$, and we must show that for every $\mathcal{A}$-structure $\mathcal{M}$ and $\alpha : X \to M$ such that $\mathcal{M} \models_\alpha \mathcal{T}$, we have $\mathcal{M} \models_\alpha \phi$. We use induction on $\mathcal{D}$.

- If $\mathcal{D}$ ends with (A), then $\phi \in \mathcal{T}$, so since $\mathcal{M} \models_\alpha \mathcal{T}$, we have $\mathcal{M} \models_\alpha \phi$.
- If $\mathcal{D}$ ends with a first-order instance of a propositional inference rule, then the same reasoning as in the proof of soundness for propositional logic (Proposition 3.27 in the notes) applies.
- If $\mathcal{D}$ ends with

  $$(=\mathrm{I})\ \frac{}{\mathcal{T} \vdash_X t = t},$$

  we have $\mathcal{M} \models_\alpha t = t \iff t^{\mathcal{M}}(\alpha) = t^{\mathcal{M}}(\alpha)$ which is clearly true.
- If $\mathcal{D}$ ends with

  $$(=\mathrm{E})\ \frac{\mathcal{T} \vdash_X s = t \quad \mathcal{T} \vdash_X \phi[x \mapsto s]}{\mathcal{T} \vdash_X \phi[x \mapsto t]}$$

  where $s, t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$ and $\phi \in \mathcal{L}^{X \cup \{x\}}_{\mathsf{form}}(\mathcal{A})$ with $\phi[x \mapsto s], \phi[x \mapsto t]$ safe, by the IH, we know

  $$\begin{aligned}
  \mathcal{M} \models_\alpha s = t &\iff s^{\mathcal{M}}(\alpha) = t^{\mathcal{M}}(\alpha), \\
  \mathcal{M} \models_\alpha \phi[x \mapsto s] &\iff \mathcal{M} \models_{\alpha\langle x \mapsto s^{\mathcal{M}}(\alpha)\rangle} \phi \quad \text{by Proposition 3.3 from Appendix} \\
  &\iff \mathcal{M} \models_{\alpha\langle x \mapsto t^{\mathcal{M}}(\alpha)\rangle} \phi \quad \text{by above} \\
  &\iff \mathcal{M} \models_\alpha \phi[x \mapsto t] \quad \text{by Proposition 3.3 again}
  \end{aligned}$$

  (where we are using, in the second line for instance, that $(x \mapsto s)^{\mathcal{M}}(\alpha) = \alpha\langle x \mapsto s^{\mathcal{M}}(\alpha)\rangle$).
- If $\mathcal{D}$ ends with

  $$(\exists\mathrm{I})\ \frac{\mathcal{T} \vdash_X \phi[x \mapsto t]}{\mathcal{T} \vdash_X \exists x\, \phi}$$

  with $\phi[x \mapsto t]$ safe, then by the IH, we have $\mathcal{M} \models_\alpha \phi[x \mapsto t]$, which (by Proposition 3.3 again) means $\mathcal{M} \models_{\alpha\langle x \mapsto t^{\mathcal{M}}(\alpha)\rangle} \phi$; thus there is $a \in M$ such that $\mathcal{M} \models_{\alpha\langle x \mapsto a\rangle} \phi$, i.e., $\mathcal{M} \models_\alpha \exists x\, \phi$.
- Finally, suppose $\mathcal{D}$ ends with

  $$(\exists\mathrm{E})\ \frac{\mathcal{T} \vdash_X \exists x\, \phi \quad \mathcal{T} \cup \{\phi\} \vdash_{X \cup \{x\}} \psi}{\mathcal{T} \vdash_X \psi}$$

  with $x \notin X$. By the first IH, we know

  $$\mathcal{M} \models_\alpha \exists x\, \phi \iff \exists a \in M \text{ s.t. } \mathcal{M} \models_{\alpha\langle x \mapsto a\rangle} \phi.$$

  Since also $\mathcal{M} \models_\alpha \mathcal{T}$ by assumption, and so $\mathcal{M} \models_{\alpha\langle x \mapsto a\rangle} \mathcal{T}$ (by Exercise 2.11) since $\mathcal{T}$ only has free variables from $X \not\ni x$, by the second IH, we get

  $$\mathcal{M} \models_{\alpha\langle x \mapsto a\rangle} \psi.$$

  Since ψ also only has free variables from $X \not\ni x$, by Exercise 2.11 again, this means

  $$\mathcal{M} \models_\alpha \psi,$$

  as desired. □

3.38. **Theorem** (completeness)**.** If $\mathcal{T} \models_X \phi$, then $\mathcal{T} \vdash_X \phi$.

Our proof strategy will be an extension of what we did in propositional logic (Theorem 3.28 in the notes). Suppose $\mathcal{T} \nvdash_X \phi$. We will show that $\mathcal{T} \not\models_X \phi$, i.e., we will construct an $\mathcal{A}$-structure $\mathcal{M}$ together with a variable assignment $\alpha : X \to M$ such that $\mathcal{M} \models_\alpha \mathcal{T}$ but $\mathcal{M} \not\models_\alpha \phi$.

As in propositional logic, we would like to define $\mathcal{M}$ to be "precisely what $\mathcal{T}$ demands". To start, we need to specify the underlying set M. Clearly, each term $t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$ will need to have an interpretation $t^{\mathcal{M}} \in M$. Thus, as a first approximation, we might take the underlying set to be simply the set of terms $\mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$, where we think of a term $t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$ as its own interpretation. However, the theory $\mathcal{T}$ requires some terms to have the same interpretation: for example,

$$\mathcal{T}_{\mathsf{abgrp}} \vdash_{\{x,y\}} x + y = y + x$$

(by a simple application of (∀E)), and so any model $\mathcal{M} \models \mathcal{T}_{\mathsf{abgrp}}$ will have to interpret $x + y, y + x$ as the same element, by soundness. We therefore take

$$M := \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})/{\equiv_{\mathcal{T}}},$$

where $\equiv_{\mathcal{T}}$ is the **$\mathcal{T}$-provable equality** relation between terms defined by

$$s \equiv_{\mathcal{T}} t \ :\Longleftrightarrow\ \mathcal{T} \vdash_X s = t.$$

In other words, M consists of elements which have to exist in any $\mathcal{A}$-structure with a variable assignment $\alpha : X \to M$, which are equal precisely when $\mathcal{T}$ says they have to be.

3.39. **Lemma.** $\equiv_{\mathcal{T}}$ is an equivalence relation on $\mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$.

Proof. By the (=I), (Sym), and (Trans) rules (Example 3.13 and Exercise 3.14). □

3.40. **Lemma.** Let $s_1, \ldots, s_n, t_1, \ldots, t_n \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$ be terms. If $s_1 \equiv_{\mathcal{T}} t_1$, …, $s_n \equiv_{\mathcal{T}} t_n$, then:

(a) For each function symbol $f \in \mathcal{A}^n_{\mathsf{fun}}$, we have $f(s_1, \ldots, s_n) \equiv_{\mathcal{T}} f(t_1, \ldots, t_n)$; thus

$$\begin{aligned} f^{\mathcal{M}} : M^n = (\mathcal{L}^X_{\mathsf{term}}(\mathcal{A})/{\equiv_{\mathcal{T}}})^n &\longrightarrow \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})/{\equiv_{\mathcal{T}}} = M \\ ([t_1], \ldots, [t_n]) &\longmapsto [f(t_1, \ldots, t_n)]. \end{aligned}$$

is a well-defined function.

(b) For each $R \in \mathcal{A}^n_{\mathsf{rel}}$, we have $\mathcal{T} \vdash_X R(s_1, \ldots, s_n) \iff \mathcal{T} \vdash_X R(t_1, \ldots, t_n)$; thus

$$\begin{aligned} R^{\mathcal{M}} : M^n = (\mathcal{L}^X_{\mathsf{term}}(\mathcal{A})/{\equiv_{\mathcal{T}}})^n &\longrightarrow \{0, 1\} \\ ([t_1], \ldots, [t_n]) &\longmapsto \begin{cases} 1 & \text{if } \mathcal{T} \vdash_X R(t_1, \ldots, t_n), \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

is well-defined.

Proof. (a) is by the (Cong) rule. The proof of (b) is similar to the derivation of (Cong) in Example 3.15: if $\mathcal{T} \vdash_X R(s_1, \ldots, s_n)$, we get $\mathcal{T} \vdash_X R(t_1, \ldots, t_n)$ by repeatedly applying (=E) with the deductions of $\mathcal{T} \vdash_X s_1 = t_1$, …, coming from $s_1 \equiv_{\mathcal{T}} t_1$, …. □

We have now defined an $\mathcal{A}$-structure $\mathcal{M}$, which we equip with the variable assignment

$$\begin{aligned} \alpha : X &\longrightarrow M = \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})/{\equiv_{\mathcal{T}}} \\ x &\longmapsto [x]. \end{aligned}$$

3.41. **Lemma.** For any $t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$, the interpretation of t in $\mathcal{M}$ is itself":

$$t^{\mathcal{M}}(\alpha) = [t].$$

Proof. By a straightforward induction on t. □

We now have the following analog of Lemma 3.32 from propositional logic, which says that truth in this $\mathcal{M}$ is exactly what $\mathcal{T}$ proves. Whereas that proof needed two conditions on $\mathcal{T}$ to make the inductive cases for $\vee, \bot$ work, here we need an additional condition for $\exists$: we define $\mathcal{T} \subseteq \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$ to

- be **consistent** if $\mathcal{T} \nvdash_X \bot$, or equivalently by ($\bot$E), $\mathcal{T} \nvdash \phi$ for at least one $\phi \in \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$;
- be **complete** if for all $\phi \in \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$, either $\mathcal{T} \vdash_X \phi$ or $\mathcal{T} \vdash_X \neg\phi$, or equivalently (as in Lemma 3.31 from propositional logic), whenever $\mathcal{T} \vdash_X \phi \vee \psi$, then $\mathcal{T} \vdash_X \phi$ or $\mathcal{T} \vdash_X \psi$;
- have the **Henkin witness property** if for all $\phi \in \mathcal{L}^{X\cup\{x\}}_{\mathsf{form}}(\mathcal{A})$, whenever $\mathcal{T} \vdash_X \exists x\, \phi$, then there is $t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$ and $\phi' \equiv_\alpha \phi$ such that $\phi'[x \mapsto t]$ is safe and $\mathcal{T} \vdash_X \phi'[x \mapsto t]$.

Note that conversely, if $\mathcal{T} \vdash_X \phi'[x \mapsto t]$, then $\mathcal{T} \vdash_X \exists x\, \phi' \equiv_\alpha \exists x\, \phi$ by ($\exists$I). Thus, the witness property can be seen as saying that $\mathcal{T}$ obeys a "converse" of ($\exists$I).

3.42. **Example.** We have $\mathcal{T}_{\mathsf{ordfield}} \vdash \exists x\, (x + x = 1)$, but there is no $\mathcal{L}_{\mathsf{ordfield}}$-term denoting $1/2$, so $\mathcal{T}_{\mathsf{ordfield}}$ does not have the witness property.

3.43. **Lemma.** Let $\mathcal{T} \subseteq \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$, and let $\mathcal{M}, \alpha$ be defined as in Lemma 3.40. Then $\mathcal{T}$ is consistent and complete and has the witness property iff for all $\phi \in \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$, we have

$$(*) \qquad \mathcal{M} \models_\alpha \phi \iff \mathcal{T} \vdash_X \phi.$$

Proof. ($\Longleftarrow$) Since $\mathcal{M} \not\models_\alpha \bot$, $\mathcal{T} \nvdash_X \bot$. For any $\phi \in \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$, either $\mathcal{M} \models_\alpha \phi$ or $\mathcal{M} \models_\alpha \neg\phi$; thus either $\mathcal{T} \vdash_X \phi$ or $\mathcal{T} \vdash_X \neg\phi$. If $\mathcal{T} \vdash_X \exists x\, \phi$, then by soundness, we know that

$$\begin{aligned}
\mathcal{M} \models_\alpha \exists x\, \phi &\iff \exists t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A}) \text{ s.t. } \mathcal{M} \models_{\alpha\langle x\mapsto[t]\rangle} \phi \\
&\iff \exists t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A}) \text{ s.t. } \mathcal{M} \models_{(x\mapsto t)^{\mathcal{M}}(\alpha)} \phi \\
&\iff \exists t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A}),\ \phi' \equiv_\alpha \phi \text{ s.t. } \phi'[x \mapsto t] \text{ is safe and } \mathcal{M} \models_{(x\mapsto t)^{\mathcal{M}}(\alpha)} \phi' \\
&\iff \exists t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A}),\ \phi' \equiv_\alpha \phi \text{ s.t. } \phi'[x \mapsto t] \text{ is safe and } \mathcal{M} \models_\alpha \phi'[x \mapsto t] \\
&\iff \exists t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A}),\ \phi' \equiv_\alpha \phi \text{ s.t. } \phi'[x \mapsto t] \text{ is safe and } \mathcal{T} \vdash_X \phi'[x \mapsto t] \quad \text{by } (*)
\end{aligned}$$

(using Corollary 4.12 and Propositions 3.3 and 4.9 from the Appendix), giving the witness property.

($\Longrightarrow$) This is mostly by induction on ϕ, except that in the $\exists x\, \phi$ case, we will need to use the IH not just for ϕ but for a substitution $\phi[x \mapsto t]$. Thus, we really need to induct on the *height* of ϕ.

- For atomic $\phi = R(t_1, \ldots, t_n)$ where $R \in \mathcal{A}^n_{\mathsf{rel}}$,

$$\begin{aligned}
\mathcal{M} \models_\alpha R(t_1, \ldots, t_n) &\iff R^{\mathcal{M}}(t_1^{\mathcal{M}}(\alpha), \ldots, t_n^{\mathcal{M}}(\alpha)) = 1 \\
&\iff R^{\mathcal{M}}([t_1], \ldots, [t_n]) = 1 \quad \text{by Lemma 3.41} \\
&\iff \mathcal{T} \vdash_X R(t_1, \ldots, t_n) \qquad \text{by definition (3.40) of } R^{\mathcal{M}}.
\end{aligned}$$

- For atomic $\phi = (s = t)$,

$$\begin{aligned}
\mathcal{M} \models_\alpha s = t &\iff [s] = s^{\mathcal{M}}(\alpha) = t^{\mathcal{M}}(\alpha) = [t] \quad \text{by Lemma 3.41} \\
&\iff s \equiv_{\mathcal{T}} t \iff \mathcal{T} \vdash_X s = t \qquad \text{by definition of quotient.}
\end{aligned}$$

- The connective cases are the same as in the proof for propositional logic (Lemma 3.32).
- Finally, suppose $(*)$ holds for all formulas of height less than that of $\exists x\, \phi$, where $\phi \in \mathcal{L}^{X\cup\{x\}}_{\mathsf{form}}(\mathcal{A})$; in particular, it holds for all safe $\phi'[x \mapsto t]$ where $t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A})$ and $\phi' \equiv_\alpha \phi$. Then as in the proof of ($\Longleftarrow$) above, we have

$$\begin{aligned}
\mathcal{M} \models_\alpha \exists x\, \phi &\iff \exists t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A}),\ \phi' \equiv_\alpha \phi \text{ s.t. } \phi'[x \mapsto t] \text{ is safe and } \mathcal{M} \models_\alpha \phi'[x \mapsto t] \\
&\iff \exists t \in \mathcal{L}^X_{\mathsf{term}}(\mathcal{A}),\ \phi' \equiv_\alpha \phi \text{ s.t. } \phi'[x \mapsto t] \text{ is safe and } \mathcal{T} \vdash_X \phi'[x \mapsto t] \text{ by IH} \\
&\iff \mathcal{T} \vdash_X \exists x\, \phi \quad \text{by } (\exists\text{I}) \text{ and the witness property for } \mathcal{T}.
\end{aligned}$$

□

Now if $\mathcal{T}$ is a complete theory with the witness property such that $\mathcal{T} \nvdash_X \phi$ (hence $\mathcal{T}$ is consistent), then by Lemma 3.43, we have $\mathcal{M} \models_\alpha \mathcal{T}$ but $\mathcal{M} \not\models_\alpha \phi$, whence $\mathcal{T} \not\models_X \phi$. So to finish the proof of the completeness theorem, we need to modify an arbitrary theory $\mathcal{T}$ to give it these properties.

- As in propositional logic, completeness will be achieved by repeatedly adding axioms to $\mathcal{T}$ until it becomes complete.
- In order to achieve the witness property, whenever $\mathcal{T} \vdash_X \exists x\, \phi$, we will add a new *variable* to X, which will serve as a witness for $\exists x\, \phi$. There will now be new formulas involving the new variable, so we will need to repeat this step (as well as the previous step) in order to fix the conditions for the new formulas.

This procedure is formalized as follows.

3.44. **Lemma.** Let $\mathcal{T} \nvdash_X \phi$. Then there is a theory $\mathcal{T}' \supseteq \mathcal{T}$ with free variables from some $X' \supseteq X$, which is complete and has the witness property (for formulas over X'), such that $\mathcal{T}' \nvdash_{X'} \phi$.

Proof of completeness theorem given Lemma 3.44. Suppose $\mathcal{T} \nvdash_X \phi$. Then by Lemma 3.44, there is $\mathcal{T}' \supseteq \mathcal{T}$ with free variables from $X' \supseteq X$, which is complete and has the witness property, such that $\mathcal{T}' \nvdash_{X'} \phi$, whence $\mathcal{T}'$ is also consistent. By Lemma 3.43, the $\mathcal{A}$-structure $\mathcal{M}$ defined above obeys

$$\mathcal{M} \models_\alpha \psi \iff \mathcal{T}' \vdash_{X'} \psi$$

for all $\psi \in \mathcal{L}^{X'}_{\mathsf{form}}(\mathcal{A})$. In particular, for all $\psi \in \mathcal{T} \subseteq \mathcal{T}'$, we have $\mathcal{T}' \vdash_{X'} \psi$ (by (A)) so $\mathcal{M} \models_\alpha \psi$ and so $\mathcal{M} \models_{\alpha|_X} \psi$ since $\mathrm{FV}(\psi) \subseteq X$ (using Exercise 2.11), i.e., $\mathcal{M} \models_{\alpha|_X} \mathcal{T}$; and since $\mathcal{T}' \nvdash_{X'} \phi$, we have $\mathcal{M} \not\models_\alpha \phi$, so again since $\mathrm{FV}(\phi) \subseteq X$, $\mathcal{M} \not\models_{\alpha|_X} \phi$. So $\mathcal{M}, \alpha|_X$ witnesses that $\mathcal{T} \not\models_X \phi$. □

To prove Lemma 3.44, we need to know: (1) we can add a single axiom to $\mathcal{T}$; (2) we can add a new *variable* to serve as a witness for an existential; and (3) we can repeat both of these steps.

3.45. **Lemma.** Let $\mathcal{T} \nvdash_X \phi$ and $\psi \in \mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$. Then either $\mathcal{T} \cup \{\psi\} \nvdash_X \phi$ or $\mathcal{T} \cup \{\neg\psi\} \nvdash_X \phi$.

Proof. As in propositional logic (Lemma 3.34). □

3.46. **Lemma.** Let $\mathcal{T} \nvdash_X \phi$, and let $\psi \in \mathcal{L}^{X\cup\{x\}}_{\mathsf{form}}(\mathcal{A})$ such that $\mathcal{T} \vdash_X \exists x\, \psi$. Then there is a variable y such that $\psi[x \mapsto y]$ is safe and $\mathcal{T} \cup \{\psi[x \mapsto y]\} \nvdash_{X\cup\{y\}} \phi$.

Proof. Let $y \notin X \cup \{x\}$ with $\psi[x \mapsto y]$ safe, so that $\exists x\, \psi \sim_\alpha \exists y\, \psi[x \mapsto y]$ (Definition 4.2 from Appendix on substitution). If we had $\mathcal{T} \cup \{\psi[x \mapsto y]\} \vdash_{X\cup\{y\}} \phi$, then by (∃E) applied to $\mathcal{T} \vdash_X \exists x\, \psi$, we would have $\mathcal{T} \vdash_X \phi$. □

3.47. **Lemma.** Let $\mathcal{T}_0 \subseteq \mathcal{T}_1 \subseteq \cdots$ be theories, with free variables from $X_0 \subseteq X_1 \subseteq \cdots$ respectively, such that $\mathcal{T}_n \nvdash_{X_n} \phi$ for each n. Then $\bigcup_n \mathcal{T}_n \nvdash_{\bigcup_n X_n} \phi$.

Proof. Suppose $\bigcup_n \mathcal{T}_n \vdash_{\bigcup_n X_n} \phi$. By syntactic compactness (Proposition 3.34), there are finitely many $\phi_1, \ldots, \phi_k \in \bigcup_n \mathcal{T}_n$ with free variables from $x_1, \ldots, x_l \in \bigcup_n X_n$ such that $\{\phi_1, \ldots, \phi_k\} \vdash_{x_1,\ldots,x_l} \phi$. Let n be large enough so that $\phi_1, \ldots, \phi_k \in \mathcal{T}_n$ and $x_1, \ldots, x_l \in X_n$. Then by variable weakening (Corollary 3.33), $\{\phi_1, \ldots, \phi_k\} \vdash_{X_n} \phi$, and then by weakening, $\mathcal{T}_n \vdash_{X_n} \phi$. □

We can now repeat step (1) to achieve completeness:

3.48. **Lemma.** Let $\mathcal{T} \nvdash_X \phi$. Then there is a complete theory $\mathcal{T}' \supseteq \mathcal{T}$, still with free variables from X, such that $\mathcal{T}' \nvdash_X \phi$.

Proof. The proof is the same as in propositional logic (Lemma 3.33): enumerate $\mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$, and for each formula, add it or its negation to $\mathcal{T}$ using Lemma 3.45, then take the union of these theories and use Lemma 3.47. (If $\mathcal{L}^X_{\mathsf{form}}(\mathcal{A})$ is uncountable, use transfinite induction or Zorn's lemma.) □

Next, we use step (2) to achieve the witness property *for all formulas with the original free variables*, after which we need to repeat both steps (1) and (2) to handle the newly added variables:

3.49. **Lemma.** Let $\mathcal{T} \nvdash_X \phi$. Then there is $\mathcal{T}' \supseteq \mathcal{T}$ with free variables from some $X' \supseteq X$ such that $\mathcal{T}' \nvdash_{X'} \phi$, and $\mathcal{T}'$ has the witness property for all existential formulas $\exists x\, \psi$ which are proved by $\mathcal{T}$ over X (rather than by $\mathcal{T}'$ over X').

Proof. The proof is similar to the previous proof, using Lemma 3.46 to extend $\mathcal{T}, X$ for each possible $\exists x\, \psi$ with free variables from X. If $\mathcal{L}_{\mathsf{form}}^{X\cup\{x\}}(\mathcal{A})$ is countable, enumerate $\mathcal{L}_{\mathsf{form}}^{X\cup\{x\}}(\mathcal{A}) = \{\psi_0, \psi_1, \dots\}$, and inductively define an increasing sequence of theories $\mathcal{T}_0 \subseteq \mathcal{T}_1 \subseteq \cdots$ with free variables from $X_0 \subseteq X_1 \subseteq \cdots \subseteq \mathcal{V}$ respectively, so that $\mathcal{T}_n \nvdash_{X_n} \phi$ for each n, as follows:

- Let $\mathcal{T}_0 := \mathcal{T}$ and $X_0 := X$.
- Given $\mathcal{T}_n$ and X_n, if $\mathcal{T}_n \nvdash_{X_n} \exists x\, \psi_n$, let $\mathcal{T}_{n+1} := \mathcal{T}_n$ and $X_{n+1} := X_n$. Otherwise, by Lemma 3.46, there is a variable y_n such that $\psi_n[x \mapsto y_n]$ is safe and
$$\mathcal{T}_{n+1} := \mathcal{T}_n \cup \{\psi_n[x \mapsto y_n]\} \nvdash_{X_{n+1} := X_n \cup \{y_n\}} \phi.$$

Now let $\mathcal{T}' := \bigcup_n \mathcal{T}_n$, $X' := \bigcup_n X_n$. Then

- $\mathcal{T}'$ has the witness property for all $\exists x\, \psi$ proved by $\mathcal{T}$ over X, since ψ must be ψ_n for some n, whence $\mathcal{T} \vdash_X \exists x\, \psi_n$ implies $\mathcal{T}_n \vdash_{X_n} \exists x\, \psi_n$ by (variable) weakening, so by definition of $\mathcal{T}_{n+1}, X_{n+1}$, we have $y_n \in X_{n+1} \subseteq X'$ and (safe) $\psi_n[x \mapsto y_n] \in \mathcal{T}_{n+1} \subseteq \mathcal{T}'$ whence $\mathcal{T}' \vdash_{X'} \psi_n[x \mapsto y_n]$ by (A).
- $\mathcal{T}' \nvdash_{X'} \phi$ by Lemma 3.47, since $\mathcal{T}_0 = \mathcal{T} \nvdash_{X_0 = X} \phi$, so by induction, $\mathcal{T}_n \nvdash_{X_n} \phi$ for each n.

If $\mathcal{L}_{\mathsf{form}}^{X\cup\{x\}}(\mathcal{A})$ is uncountable, use either transfinite induction or Zorn's lemma. □

Proof of Lemma 3.44. Define an increasing sequence of theories $\mathcal{T}_0 \subseteq \mathcal{T}_1 \subseteq \cdots$, with free variables from $X_0 \subseteq X_1 \subseteq \cdots \subseteq \mathcal{V}$, so that $\mathcal{T}_n \nvdash_{X_n} \phi$ for each n, by induction as follows:

- Let $\mathcal{T}_0 := \mathcal{T}$ and $X_0 := X$.
- Given $\mathcal{T}_n$ and X_n, by Lemma 3.48, there is complete $\mathcal{T}_n' \supseteq \mathcal{T}_n$ with free variables from X_n such that $\mathcal{T}_n' \nvdash_{X_n} \phi$, and then by Lemma 3.49, there is $\mathcal{T}_{n+1} \supseteq \mathcal{T}_n'$ with free variables from $X_{n+1} \supseteq X_n$ which has the witness property for all existentials proved by $\mathcal{T}_n'$ over X_n and still satisfies $\mathcal{T}_{n+1} \nvdash_{X_{n+1}} \phi$.

Let $\mathcal{T}' := \bigcup_n \mathcal{T}_n$, $X' := \bigcup_n X_n$. Then

- $\mathcal{T}'$ is complete (over X'), since for any $\psi \in \mathcal{L}_{\mathsf{form}}^{X'}(\mathcal{A})$, we have $\mathrm{FV}(\psi) \subseteq X_n$ for some n, whence by completeness of $\mathcal{T}_n'$, either $\mathcal{T}_n' \vdash_{X_n} \psi$ or $\mathcal{T}_n' \vdash_{X_n} \neg\psi$, and so by (variable) weakening, either $\mathcal{T}' \vdash_{X'} \psi$ or $\mathcal{T}' \vdash_{X'} \neg\psi$.
- $\mathcal{T}'$ has the witness property (over X'), since for any $\exists x\, \psi \in \mathcal{L}_{\mathsf{form}}^{X'}(\mathcal{A})$ such that $\mathcal{T}' \vdash_{X'} \exists x\, \psi$, by syntactic compactness, we have $\mathcal{T}_n \vdash_{X_n} \exists x\, \psi$ for some n, whence $\mathcal{T}_n' \vdash_{X_n} \exists x\, \psi$ by weakening, so since $\mathcal{T}_{n+1}$ has the witness property for existentials proved by $\mathcal{T}_n'$ over X_n, there is $t \in \mathcal{L}_{\mathsf{term}}^{X_{n+1}}(\mathcal{A})$ and $\psi' \equiv_\alpha \psi$ with $\psi'[x \mapsto t]$ safe and $\mathcal{T}_{n+1} \vdash_{X_{n+1}} \psi'[x \mapsto t]$, so by (variable) weakening, $\mathcal{T}' \vdash_{X'} \psi'[x \mapsto t]$.
- $\mathcal{T}' \nvdash_{X'} \phi$ by Lemma 3.47.

(Note that there is no need to go into the transfinite here, since we're not enumerating formulas.) □

This concludes the proof of the completeness theorem.

3.50. **Corollary** (of soundness and completeness)**.** For $\mathcal{T} \subseteq \mathcal{L}_{\mathsf{form}}^X(\mathcal{A})$ and $\phi \in \mathcal{L}_{\mathsf{form}}^X(\mathcal{A})$,
$$\mathcal{T} \vdash_X \phi \iff \mathcal{T} \models_X \phi.$$

3.51. **Corollary.** $\mathcal{T}$ is consistent iff it is satisfiable.

4. Compactness

By syntactic compactness (Proposition 3.34) and Corollary 3.50,

4.1. **Corollary** (compactness)**.** If $\mathcal{T} \models_X \phi$, then there are finite $\mathcal{T}' \subseteq \mathcal{T}$ and $X' \subseteq X$ containing all free variables in $\mathcal{T}'$ such that $\mathcal{T}' \models_{X'} \phi$.

4.2. **Corollary.** If $\mathcal{T}$ is a theory with free variables from X, and every finite $\mathcal{T}' \subseteq \mathcal{T}$ with free variables from finite $X' \subseteq X$ is satisfiable (over X', i.e., there is a structure $\mathcal{M}$ and $\alpha : X' \to M$ with $\mathcal{M} \models_\alpha \mathcal{T}'$), then $\mathcal{T}$ is satisfiable (over X, i.e., there is $\mathcal{M}$ and $\alpha : X \to M$ with $\mathcal{M} \models_\alpha \mathcal{T}$).

Proof. Take $\phi := \bot$. □

As a first application, we have a general phenomenon of first-order logic: the inability to tell infinite cardinalities apart.

4.3. **Theorem** (upward Löwenheim–Skolem[5])**.** Let $\mathcal{T} \subseteq \mathcal{L}^{\varnothing}_{\mathsf{form}}(\mathcal{A})$ have models of cardinality $\geq n$ for each $n \in \mathbb{N}$, i.e.,

(i) either $\mathcal{T}$ has an infinite model,
(ii) or $\mathcal{T}$ has arbitrarily large finite models.

Then $\mathcal{T}$ has models of cardinality $\geq |X|$ for every set X.

Proof. As usual, we treat X as a set of variables. Consider

$$\mathcal{T}' := \mathcal{T} \cup \{\neg(x = y) \mid x \neq y \in X\}$$

with free variables from X. A model of it consists of an $\mathcal{A}$-structure $\mathcal{M}$ together with a variable assignment $\alpha : X \to M$ such that $\mathcal{M} \models_\alpha \mathcal{T}'$, which means $\mathcal{M} \models \mathcal{T}$ (since $\mathcal{T}$ has no free variables), and also $\mathcal{M} \models_\alpha \neg(x = y)$ for each $x \neq y \in X$, which means exactly that $\alpha : X \to M$ is injective. So $\mathcal{T}'$ is satisfiable iff $\mathcal{M}$ has a model of cardinality $\geq |X|$. By compactness, it suffices to show that for every finite $X' \subseteq X$ and $\mathcal{T}'' \subseteq \mathcal{T}'$ with free variables from X', there is a model $\mathcal{M}' \models_{\alpha'} \mathcal{T}''$ where $\alpha' : X' \to M'$. Indeed, since $|X'| \in \mathbb{N}$, we may let $\mathcal{M}'$ be a model of cardinality $\geq |X'|$, and $\alpha' : X' \to M'$ be an injection, so that $\mathcal{M}' \models_{\alpha'} \mathcal{T}''$ for the same reason as before. □

4.4. **Example.** The following classes $\mathcal{K}$ of structures are *not* axiomatizable in first-order logic:

- all finite sets
- all finite fields
- all finite abelian groups
- all finite posets
- all finite graphs
- ...

Indeed, since there are arbitrarily large finite structures of each of these types, any theory satisfied by all of them must also be satisfied by some infinite structure, by the preceding Theorem.

4.5. **Example.** For any infinite $\mathcal{A}$-structure $\mathcal{M}$, the class of structures isomorphic to $\mathcal{M}$ is *not* first-order axiomatizable, since any theory satisfied by $\mathcal{M}$ must also be satisfied by a structure of cardinality $> |M|$ (e.g., $\geq |\mathcal{P}(M)|$, which implies $> |M|$ by Cantor's theorem).

[5]This is one formulation of the theorem that usually goes by this name. Other versions may be obtained by applying this version to the theory of a particular infinite structure in a language that includes a constant symbol for each element, and/or by carefully counting the cardinalities of the models constructed in the proof of the completeness theorem above.

4.6. **Example.** Consider the structure $\mathcal{M} := \mathbb{N}$ equipped with the usual linear order $\leq$. By Löwenheim–Skolem, there is another $\{\leq\}$-structure $\mathcal{N}$ that is uncountable but also satisfies $\mathrm{Th}(\mathbb{N})$. This means $\mathcal{N}$ "looks just like" $\mathbb{N}$, as far as the first-order properties of the ordering can tell, despite being much larger:

- In particular, since $\mathbb{N}$ is linearly ordered, $\mathrm{Th}(\mathbb{N})$ includes the linear order axioms (Example 2.25), hence $\mathcal{N}$ is also linearly ordered.
- Since $\mathbb{N}$ has a least element, i.e., satisfies the axiom $\exists x_0\, \forall y\, (x_0 \leq y)$, so does $\mathcal{N}$; let us call its least element $0_{\mathcal{N}}$ (to distinguish it from the number 0).
- Since $\mathbb{N}$ has a second-least element, i.e., satisfies the axiom
$$\exists x_1\, \exists x_0\, (\neg(x_0 = x_1) \wedge \forall y\, (\neg(x_0 = y) \leftrightarrow x_1 \leq y)),$$
so does $\mathcal{N}$; let us call its least element $1_{\mathcal{N}}$.
- Similarly, $\mathcal{N}$ has a next-least element $2_{\mathcal{N}}$, followed by $3_{\mathcal{N}}$, etc.
- However, since $\mathcal{N}$ is uncountable, it must contain other elements which are strictly greater than $n_{\mathcal{N}}$ for each $n \in \mathbb{N}$. In other words, $\mathcal{N}$ must contain "infinite" elements (even though it looks just like $\mathbb{N}$)!

4.7. **Exercise.** Show that $\mathcal{N}$ above can be partitioned into the initial segment $\{0_{\mathcal{N}}, 1_{\mathcal{N}}, \dots\}$ (which is order-isomorphic to $\mathbb{N}$), followed by some uncountable number of contiguous intervals each of which is order-isomorphic to $\mathbb{Z}$.

(In fact it is possible to construct explicit examples of such $\mathcal{N}$, e.g., $\mathbb{N}$ followed by $\mathbb{R} \times \mathbb{Z}$ ordered lexicographically. However, it takes some work to show that this linear order indeed satisfies $\mathrm{Th}(\mathbb{N})$.)

4.8. **Example.** Consider the ordered field $\mathbb{R}$. By Löwenheim–Skolem, there is another ordered field $\mathcal{R}$ that has much larger cardinality than $\mathbb{R}$ (e.g., $\geq |\mathcal{P}(\mathbb{R})|$), but also satisfies $\mathrm{Th}(\mathbb{R})$. Similarly to the preceding example, it can be shown that any such field must contain "infinite elements", i.e., $\geq n := 1 + \cdots + 1$ for all $n \in \mathbb{N}$, and hence (being an ordered field) also "infinitesimals" which are positive but $\leq 1/n$ for every $n > 0$. In other words, there are "number systems" which look just like the reals, as far as first-order logic can tell, but which have infinities!

4.9. **Remark.** In fact, $\mathrm{Th}(\mathbb{R})$ has been extensively studied, and its models are well-understood. A set of axioms implying all of $\mathrm{Th}(\mathbb{R})$ consists of the ordered field axioms (Example 2.31), plus

- every positive element has a square root;
- every odd-degree polynomial has a root (an axiom schema, one for each odd degree n saying $\forall a_0 \cdots \forall a_{n-1}\, \exists x\, (x^n + a_{n-1}x^{n-1} + \cdots + a_0 = 0)$, where x^n is an abbreviation for $x \cdots x$).

Models of these axioms are known as **real-closed fields**.[6] There are many interesting examples, the smallest one being the field of real algebraic numbers (i.e., real roots of rational polynomials, such as $\sqrt{2}$ or $\sqrt[3]{\sqrt{5} - \sqrt{7}}$) which is countable. Examples larger than $\mathbb{R}$ may be built by considering certain fields of power series like $x^2 - x^{\frac{1}{2}} + x^{\frac{1}{3}} - \cdots$.

4.10. **Exercise.** Apply the Löwenheim–Skolem theorem to the infinite line graph on $\mathbb{Z}$:

$$\cdots \text{ —— } -2 \text{ —— } -1 \text{ —— } 0 \text{ —— } 1 \text{ —— } 2 \text{ —— } 3 \text{ —— } \cdots$$

What conclusions can you draw from an uncountable model of $\mathrm{Th}(\mathbb{Z})$?

4.11. **Exercise.** Show that the class of **locally finite graphs**, i.e., graphs in which each vertex has finitely many neighbors, is not axiomatizable by a first-order theory.

4.12. **Exercise.** Show that the class of **well-orders**, i.e., linear orders in which there is no infinite decreasing sequence, is not axiomatizable by a first-order theory.

[6] These are to $\mathbb{R}$ as algebraically closed fields of characteristic 0 are to $\mathbb{C}$.

4.A. **Application: nonstandard analysis.** We now give a generalization of the Löwenheim–Skolem Theorem 4.3, that provides a ready-made recipe for producing structures like those above that "look like" a familiar set, but with extra elements in "every place they could possibly appear".

4.13. **Definition.** Let M be an arbitrary set. The **complete signature** of M is the first-order signature $\mathcal{A}_M$ with a symbol denoting *every* possible function or relation on M. That is, for $n \in \mathbb{N}$,

$$(\mathcal{A}_M)^n_{\mathsf{fun}} := M^{M^n}, \qquad (\mathcal{A}_M)^n_{\mathsf{rel}} := \mathcal{P}(M^n).$$

The **complete structure** of M is the $\mathcal{A}_M$-structure $\mathcal{M}$ with each symbol interpreted as itself:

$$f^{\mathcal{M}} := f : M^n \to M, \qquad R^{\mathcal{M}} := R \subseteq M^n$$

for each $f \in (\mathcal{A}_M)^n_{\mathsf{fun}}$ and $R \in (\mathcal{A}_M)^n_{\mathsf{rel}}$.

4.14. **Theorem.** For any set M, there is another model ${}^*\mathcal{M}$ of the $\mathcal{A}_M$-theory $\mathrm{Th}(\mathcal{M})$ of the complete structure of M, called a **nonstandard extension** of M, which "contains elements satisfying any finitely consistent family of conditions", in the following sense. Let us write

$${}^*f := f^{{}^*\mathcal{M}} : {}^*M^n \to {}^*M, \qquad {}^*R := R^{{}^*\mathcal{M}} \subseteq {}^*M^n$$

for the interpretations in ${}^*\mathcal{M}$ of $f : M^n \to M$ and $R \subseteq M^n$. Then for any $n \in \mathbb{N}$ and family of n-ary relations $\mathcal{F} \subseteq \mathcal{P}(M^n)$ with the **finite intersection property**, meaning $R_1 \cap \cdots \cap R_k \neq \varnothing$ for any finitely many $R_1, \ldots, R_k \in \mathcal{F}$, we have $\bigcap_{R\in\mathcal{F}} {}^*R \neq \varnothing$.

Proof. For each $\mathcal{F} \subseteq \mathcal{P}(M^n)$ with the finite intersection property, let $x_{\mathcal{F},1}, \ldots, x_{\mathcal{F},n}$ be new variables, and let X be the set of all of these new variables, for all $\mathcal{F}$. Consider the $\mathcal{A}_M$-theory

$$\mathcal{T} := \mathrm{Th}(\mathcal{M}) \cup \{R(x_{\mathcal{F},1}, \ldots, x_{\mathcal{F},n}) \mid \mathcal{F} \subseteq \mathcal{P}(M^n) \text{ has FIP}, R \in \mathcal{F}\} \subseteq \mathcal{L}^X_{\mathsf{form}}(\mathcal{A}_M).$$

A model consists of ${}^*\mathcal{M} \models \mathrm{Th}(\mathcal{M})$ equipped with a variable assignment $\alpha : X \to {}^*M$ such that

$$(\alpha(x_{\mathcal{F},1}), \ldots, \alpha(x_{\mathcal{F},n})) \in \bigcap_{R\in\mathcal{F}} {}^*R$$

for each family $\mathcal{F}$ with the finite intersection property. Thus by compactness, it suffices to show that for every finite $X' \subseteq X$ and finite $\mathcal{T}' \subseteq \mathcal{T}$ with free variables from X', $\mathcal{T}'$ is satisfiable. We may assume without loss of generality that X' consists of $x_{\mathcal{F},1}, \ldots, x_{\mathcal{F},n}$ for all $\mathcal{F}$ in some finite list $\mathcal{F}_1, \ldots, \mathcal{F}_m$ each with the FIP (by enlarging X' to such a set if necessary). Then each axiom in $\mathcal{T}'$ is either in $\mathrm{Th}(\mathcal{M})$, or of the form $R(x_{\mathcal{F},1}, \ldots, x_{\mathcal{F},n})$ for some R in one of $\mathcal{F} = \mathcal{F}_1, \ldots, \mathcal{F}_m$. Define

$$\begin{aligned} \alpha : X' &\longrightarrow M \\ x_{\mathcal{F},1}, \ldots, x_{\mathcal{F},n} &\longmapsto \text{any } n\text{-tuple in } \bigcap_{\substack{R\in\mathcal{F} \\ R(x_{\mathcal{F},1},\ldots,x_{\mathcal{F},n})\in\mathcal{T}'}} R \end{aligned}$$

for each $\mathcal{F} = \mathcal{F}_1, \ldots, \mathcal{F}_m$, using that each such intersection is nonempty by the FIP and the fact that $\mathcal{T}'$ is finite. Then $\mathcal{M} \models_\alpha \mathcal{T}'$. □

4.15. **Remark.** In particular, in the above setup, for each element $a \in M$, we may treat a as a constant (nullary function); then a is denoted by a constant symbol in $\mathcal{A}_M$, hence also has an interpretation ${}^*a = a^{{}^*\mathcal{M}}$ in the $\mathcal{A}_M$-structure ${}^*\mathcal{M}$. This defines a function

$$\begin{aligned} {}^* : M &\longrightarrow {}^*M \\ a &\longmapsto {}^*a. \end{aligned}$$

This function is injective, since for any $a \neq b \in M$, the complete structure $\mathcal{M}$ satisfies the $\mathcal{A}_M$-sentence $\neg(a = b)$, which is hence in $\mathrm{Th}(\mathcal{M})$, hence also satisfied by ${}^*M$. Hence, we may think of the function * as "embedding" M inside the (usually much larger) ${}^*M$.

4.16. **Example.** Consider $M = \mathbb{N}$. The above gives us a bigger set ${}^*\mathbb{N}$, together with an injection

$$^* : \mathbb{N} \hookrightarrow {}^*\mathbb{N}.$$

Moreover, ${}^*\mathbb{N}$ has its own versions ${}^*f, {}^*R$ of every function f or relation R on $\mathbb{N}$; and these all behave the same way (as far as first-order logic can tell) as in $\mathbb{N}$.

For example, the linear order $\leq \subseteq \mathbb{N}^2$ extends to a binary relation ${}^*\leq \subseteq {}^*\mathbb{N}^2$, which is also a linear order because $\leq$ was; and $^* : \mathbb{N} \to {}^*\mathbb{N}$ is an order-embedding, i.e., for any $a, b \in \mathbb{N}$,

$$a \leq b \iff {}^*a \; {}^*\leq {}^*b,$$

since if $a \leq b$ holds in $\mathbb{N}$ then "$a \leq b$" is an $\mathcal{A}_\mathbb{N}$-sentence in $\mathrm{Th}(\mathbb{N})$ hence also satisfied by ${}^*\mathbb{N}$, and similarly if $a \not\leq b$ in $\mathbb{N}$ then "$\neg(a \leq b)$" is satisfied by ${}^*\mathbb{N}$. Moreover, ${}^*0 \in {}^*\mathbb{N}$ is the least element, ${}^*1 \in {}^*\mathbb{N}$ is the second-least, etc., again since these are described by first-order $\mathcal{A}_\mathbb{N}$-sentences.

We also have the unary relations $U_n := \{a \in \mathbb{N} \mid a \geq n\} \subseteq \mathbb{N}$; hence we get unary relations ${}^*U_n \subseteq {}^*\mathbb{N}$. The family $\{U_n\}_{n\in\mathbb{N}}$ satisfies the FIP, since $U_m \cap U_n = U_{\max(m,n)}$; hence

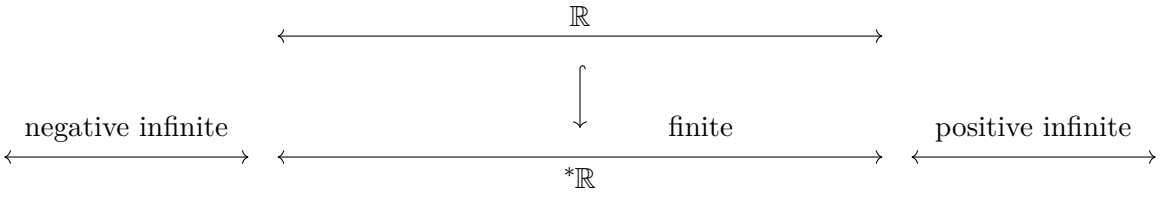
$$\bigcap_{n\in\mathbb{N}} {}^*U_n \neq \varnothing.$$

We have ${}^*U_n = \{a \in {}^*\mathbb{N} \mid a \; {}^*\geq {}^*n\}$, i.e., $\forall a \in {}^*\mathbb{N}\,(a \in {}^*U_n \iff a \; {}^*\geq {}^*n)$, since the $\mathcal{A}_\mathbb{N}$-sentence $\forall x\,(U_n(x) \leftrightarrow x \geq n)$ (where n is a constant symbol) holds in $\mathbb{N}$, hence also in ${}^*\mathbb{N}$. Thus an element of $\bigcap_{n\in\mathbb{N}} {}^*U_n$ is "infinite", i.e., strictly bigger than ${}^*n$ for each $n \in \mathbb{N}$, thereby recovering Example 4.6.

Unlike that example, here we also have every other first-order structure we can dream of on $\mathbb{N}$. For example, there is a binary operation ${}^*+ : {}^*\mathbb{N}^2 \to {}^*\mathbb{N}$, which is associative, commutative, has identity element ^{*0}, order-preserving, etc. Similarly there is a binary operation ${}^*\cdot : {}^*\mathbb{N}^2 \to {}^*\mathbb{N}$ which distributes over ${}^*+$, etc. The set of prime numbers $P \subseteq \mathbb{N}$ is a unary relation, hence we get a set of "nonstandard primes" ${}^*P \subseteq {}^*\mathbb{N}$, such that $a \in {}^*P$ iff $a \; {}^*\geq {}^*2$ and a cannot be written as $b \; {}^*\cdot \; c$ for any $b, c \; {}^*\geq {}^*2$, etc. Note that ${}^*P$ also contains "infinite primes", since $P \cap U_n \neq \varnothing$ for each n.

4.17. **Exercise.** Show that the twin primes conjecture holds iff ${}^*P$ contains two infinite elements which differ by ^{*2}.

4.18. **Example.** Similarly, we may consider $M = \mathbb{R}$, to get a nonstandard extension ${}^*\mathbb{R}$ of $\mathbb{R}$, which is an ordered field just like $\mathbb{R}$. As in Example 4.16, there are elements $a \in {}^*\mathbb{R}$ which are $a \; {}^*> {}^*r$ for any $r \in \mathbb{R}$; we call such elements **positive infinite**. Similarly, we call $b \in {}^*\mathbb{R}$ **negative infinite** if $b \; {}^*< {}^*s$ for any $s \in \mathbb{R}$. If $c \in {}^*\mathbb{R}$ is neither positive infinite nor negative infinite, then that means ${}^*s \; {}^*\leq c \; {}^*\leq {}^*r$ for some $r, s \in \mathbb{R}$; we call such c **finite**. Thus the ${}^*\mathbb{R}$-line looks like:



Note that there is no smallest positive infinite $a \in {}^*\mathbb{R}$, since if a is positive infinite, then so is $a \; {}^*- \; {}^*1$ (since for any $r \in \mathbb{R}$, we have ${}^*r \; {}^*+ \; {}^*1 = {}^*(r+1) \; {}^*< a$, hence ${}^*r \; {}^*< a \; {}^*- \; {}^*1$).

Note also that the "finite" part of the ${}^*\mathbb{R}$-line contains much more than the image of $^* : \mathbb{R} \to {}^*\mathbb{R}$. Indeed, just as positive infinite elements exist, we also have

$$\bigcap_{0<\varepsilon\in\mathbb{R}} {}^*(0, \varepsilon) \neq \varnothing$$

since the family of sets $\{(0, \varepsilon)\}$ has the FIP in $\mathbb{R}$; elements of this set are called **positive infinitesimal**. Similarly, negative elements which are $a \; {}^*> {}^*-\varepsilon$ for every $\varepsilon > 0$ are **negative infinitesimal**. More generally, we call $c \in {}^*\mathbb{R}$ **infinitesimal** if it is $c \; {}^*< {}^*\varepsilon$ and $c \; {}^*> {}^*-\varepsilon$ for every positive real ε.

4.19. **Exercise.**

(a) Show that $c \in {}^*\mathbb{R}$ is positive infinitesimal iff ${}^*1 {}^*/ c$ is positive infinite.

(b) Show that

$$\begin{aligned} \text{finite } {}^*+ \text{ finite} &= \text{finite}, \\ \text{finite } {}^*\cdot \text{ finite} &= \text{finite}, \\ \text{infinitesimal } {}^*+ \text{ infinitesimal} &= \text{infinitesimal}, \\ \text{finite } {}^*\cdot \text{ infinitesimal} &= \text{infinitesimal}, \end{aligned}$$

Thus, the set of finite elements forms a subring of ${}^*\mathbb{R}$, inside of which the infinitesimal elements form an ideal. In particular,

$$a \approx b \ :\iff \ a \,{}^*- b \text{ is infinitesimal}$$

defines a (coset) equivalence relation on ${}^*\mathbb{R}$.

(c) Show that every finite $c \in {}^*\mathbb{R}$ is $\approx {}^*r$ for a unique $r \in \mathbb{R}$, called the **standard part** of c. [Hint: take a supremum.]

(d) Conclude that the quotient ring {finite elements of ${}^*\mathbb{R}$}/{infinitesimals} is isomorphic to $\mathbb{R}$.

You probably know that the early pioneers of calculus thought in terms of "infinitesimals", e.g., the derivative of a function f at x is given by

$$f'(x) = \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

where Δx is a nonzero but "infinitesimally small" real number. This wasn't made rigorous until the ε-δ definition was introduced, which did away entirely with infinitesimals and infinities, except as an intuitive motivating concept.

Much later, in the 1960s, Robinson showed that the "infinitesimal" view can in fact also be made rigorous, with the help of first-order logic: the key insight is that the "infinitesimal" Δx is *not* a real number, but rather lives in the extended field ${}^*\mathbb{R}$, which however (being a model of $\mathrm{Th}(\mathbb{R})$) obeys the same first-order properties as $\mathbb{R}$. The resulting viewpoint is called **nonstandard analysis**, and has since found widespread applications in many areas, not just analysis. Here is a small sampler:

4.20. **Proposition.** Let $f : \mathbb{R} \to \mathbb{R}$ be a function and $a, b \in \mathbb{R}$. We have

$$\lim_{x \to a} f(x) = b$$

iff for all infinitesimal ${}^*0 \neq \Delta a \in {}^*\mathbb{R}$, we have ${}^*f({}^*a \,{}^*+ \Delta a) \approx {}^*b$.

Proof. The standard definition of $\lim_{x \to a} f(x) = b$ is

$$(*) \qquad \forall \varepsilon > 0 \ \exists \delta > 0 \ \forall -\delta < \Delta a < \delta, \ \Delta a \neq 0 \ (-\varepsilon < f(a + \Delta a) - b < \varepsilon).$$

For fixed $\varepsilon > 0$ and $\delta > 0$, the statement

$$\forall -\delta < \Delta a < \delta, \ \Delta a \neq 0 \ (-\varepsilon < f(a + \Delta a) - b < \varepsilon)$$

is the interpretation in $\mathbb{R}$ of a first-order $\mathcal{A}_{\mathbb{R}}$-sentence (where every symbol except for the variable Δa is part of the signature $\mathcal{A}_{\mathbb{R}}$). Thus, it is equivalent to the interpretation in ${}^*\mathbb{R}$:

$$\forall {}^*-\delta \,{}^*< \Delta a \,{}^*< {}^*\delta, \ \Delta a \neq {}^*0 \ ({}^*-\varepsilon \,{}^*< {}^*f({}^*a \,{}^*+ \Delta a) \,{}^*- {}^*b \,{}^*< {}^*\varepsilon).$$

Thus, $(*)$ is equivalent to

$$(\dagger) \qquad \forall \varepsilon > 0 \ \exists \delta > 0 \ \forall {}^*-\delta \,{}^*< \Delta a \,{}^*< {}^*\delta, \ \Delta a \neq {}^*0 \ ({}^*-\varepsilon \,{}^*< {}^*f({}^*a \,{}^*+ \Delta a) \,{}^*- {}^*b \,{}^*< {}^*\varepsilon).$$

(Note that we did not translate the *entire* $(*)$ to ${}^*\mathbb{R}$, but only the part inside the first two quantifiers.)

If (†) holds, then for every infinitesimal ${}^*0 \neq \Delta a \in {}^*\mathbb{R}$, for every $\varepsilon > 0$, letting $\delta > 0$ be given by (†), we have that ${}^*{-}\delta \mathrel{{}^*{<}} \Delta a \mathrel{{}^*{<}} {}^*\delta$ (since Δa is infinitesimal), thus by (†),

$$ {}^*{-}\varepsilon \mathrel{{}^*{<}} {}^*f({}^*a {}^*{+} \Delta a) {}^*{-} {}^*b \mathrel{{}^*{<}} {}^*\varepsilon; $$

since $\varepsilon > 0$ was arbitrary, this shows that ${}^*f({}^*a {}^*{+} \Delta a) {}^*{-} {}^*b$ is infinitesimal, i.e., ${}^*f({}^*a {}^*{+} \Delta a) \approx {}^*b$.

Conversely, if $(*)$ fails, there is some $\varepsilon > 0$ such that for every $\delta > 0$,

$$ D_\delta := \{\Delta a \in \mathbb{R} \mid (-\delta < \Delta a < \delta) \wedge (\Delta a \neq 0) \wedge \neg(-\varepsilon < f(a + \Delta a) - b < \varepsilon)\} \neq \varnothing. $$

Note that for $\delta \leq \delta'$, we have $D_\delta \subseteq D_{\delta'}$; thus these sets D_δ have the finite intersection property. So there is $\Delta a \in \bigcap_{\delta>0} {}^*D_\delta$, which by reinterpreting the formula defining D_δ in ${}^*\mathbb{R}$ means

$$ \forall \delta > 0\, ({}^*{-}\delta \mathrel{{}^*{<}} \Delta a \mathrel{{}^*{<}} {}^*\delta) \wedge \Delta a \neq {}^*0 \wedge \neg({}^*{-}\varepsilon \mathrel{{}^*{<}} {}^*f({}^*a {}^*{+} \Delta a) {}^*{-} {}^*b \mathrel{{}^*{<}} {}^*\varepsilon), $$

i.e., Δa is nonzero infinitesimal but ${}^*f({}^*a {}^*{+} \Delta a) {}^*{-} {}^*b$ is not infinitesimal, i.e., ${}^*f({}^*a {}^*{+} \Delta a) \not\approx {}^*b$. □

4.21. **Exercise.** Using the preceding result and Exercise 4.19, give simple proofs of the limit laws:

$$ \lim_{x\to a}(f(x) + g(x)) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x), $$
$$ \lim_{x\to a}(f(x) \cdot g(x)) = \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x), $$

assuming the limits of f, g exist.

It should be clear from Proposition 4.20 that the translation from the classical ε-δ definition to the nonstandard one is completely mechanical, and has nothing to do with limits *per se*. Similarly:

4.22. **Exercise.** Let $f : \mathbb{R} \to \mathbb{R}$.

(a) For $b \in \mathbb{R}$, $\lim_{x\to+\infty} f(x) = b$ iff for all positive infinite $a \in {}^*\mathbb{R}$, we have ${}^*f(a) \approx {}^*b$.

(b) f is continuous iff

$$ \forall a \in \mathbb{R}\, \forall b \in {}^*\mathbb{R}\, ({}^*a \approx b \implies {}^*f({}^*a) \approx {}^*f(b)). $$

(c) f is *uniformly* continuous iff

$$ \forall a \in {}^*\mathbb{R}\, \forall b \in {}^*\mathbb{R}\, (a \approx b \implies {}^*f(a) \approx {}^*f(b)). $$

Recall that "limits of truth assignments" also played an important role in propositional logic (Definition 2.57 in notes). We can similarly formulate this notion of "limit" in nonstandard terms:

4.23. **Exercise.** Let $\mathcal{A}$ be an alphabet (for propositional logic), let $M := \{0,1\}^{\mathcal{A}}$ be the set of truth assignments, and let ${}^*M$ be a nonstandard extension. For $a, b \in {}^*M$, we write

$$ a \approx b \;:\Longleftrightarrow\; \forall \phi \in \mathcal{L}(\mathcal{A})\, (a \in {}^*\mathrm{Mod}(\phi) \implies b \in {}^*\mathrm{Mod}(\phi)), $$

pronounced "a, b are **infinitesimally close**". (Note: we do not define "infinitesimal" here.)

(a) Show that $\approx$ is a symmetric, hence an equivalence relation on ${}^*M$.

(b) Show that every $a \in {}^*M$ is $\approx {}^*m$ for a unique $m \in M$, called the **standard part** of a.

(c) Show that a set of truth assignments $\mathcal{K} \subseteq M$ is axiomatizable iff every $m \in M$ for which ${}^*m$ is infinitesimally close to some $a \in {}^*\mathcal{K}$ is in $\mathcal{K}$.

(d) Show that a set of truth assignments $\mathcal{K} \subseteq M$ is axiomatizable by a single formula iff for every two $a, b \in {}^*M$ which are infinitesimally close, we have $a \in {}^*\mathcal{K}$ iff $b \in {}^*\mathcal{K}$.