CS 221 **Programming Languages** **Handout 1**
Fall 2011 **Sept 29, 2011**

Induction Tutorial

1 Basic concepts

Defn 1 (poset) *A **poset** is a structure $\langle A, \leq \rangle$ where A is a set and $\leq$ is a binary relation on A, such that $\leq$ is reflexive, antisymmetric, and transitive. (Usual derivative relations: $<$, $\geq$, $>$.)*

Defn 2 (total order) *A **poset** is **totally ordered** (*linearly ordered*, a* chain*) if any two elements are comparable, i. e., given two elements a and b, either $a \leq b$ or $b \leq a$.*

Defn 3 (well founded) *A poset $\langle A, \leq \rangle$ is **well founded** if every non-empty subset of A contains a minimal element.*

Defn 4 (DCC) *A poset $\langle A, \leq \rangle$ satisfies the **Decending Chain Condition** (DCC) if there does not exist an infinite decending chain of elements of A, i. e., an infinite sequence $a_0 > a_1 > a_2 > \cdots$ where each successive element is strictly less than its predecessor.*

Prop 1 *A poset $\langle A, \leq \rangle$ is well founded iff it satisfies the DCC.*

Defn 5 (well ordered) *A poset $\langle A, \leq \rangle$ is **well ordered** if every non-empty subset of A contains a* least *(or* minimum*) element.*

Prop 2 *A poset is well ordered iff it is well founded and totally ordered.*

Examples: well founded posets

1. $\langle \mathbb{N}, \leq \rangle$, the natural numbers with the usual ordering
2. $\langle \mathcal{P}(\{0, \ldots, 100\}), \subseteq \rangle$, subsets of the set $\{0, \ldots, 100\}$, ordered by the subset relation.
3. $\langle \mathbb{N} \times \mathbb{N}, \leq_L \rangle$, where $\leq_L$ is the *lexicographical* ordering of pairs of numbers: $(a_1, b_1) \leq_L (a_2, b_2)$ iff $a_1 < a_2$ or $a_1 = a_2$ and $b_1 \leq b_2$.
4. $\langle \mathbb{N} \times \mathbb{N}, \leq_P \rangle$, where $\leq_P$ is the *pointwise* ordering of pairs of numbers: $(a_1, b_1) \leq_P (a_2, b_2)$ iff $a_1 \leq a_2$ and $b_1 \leq b_2$.
5. $\langle \mathbb{N} \cup \{\infty\}, \leq \rangle$, the natural numbers extended with an infinite element ∞, where $n < \infty$ for all $n \in \mathbb{N}$ (this is the same as the ordinal number $\omega + 1$).

Examples: non-well-founded posets

1. $\langle \mathbb{N}, \geq \rangle$, the natural numbers with the inverse ordering.
2. $\langle \mathcal{P}(\mathbb{N}), \subseteq \rangle$
3. $\langle \mathcal{P}(S), \subseteq \rangle$ for any infinite set S.

Exercise 1 *Let A be the aphabetic characters $\{a, \ldots, z\}$ with the usual ordering, and let A^* be the set of finite strings of characters. For $s, t \in A$, define $s \leq t$ as the usual lexicographic (i. e.dictionary) ordering of strings. Is $\langle A^*, \leq \rangle$ a well founded set?*

Exercise 2 *Give an infinite set S, show how to construct an inifinite descending chain of subsets of S.*

2 Well Founded Induction

The *Principle of Well Founded Induction* (WFI) is defined as follows. Given a well founded poset $\langle A, \leq \rangle$, and a unary relation P on A (in other words, a subset of A),

$$\forall x.((\forall y.\ y < x \Rightarrow P(y)) \Rightarrow P(x)) \Rightarrow \forall z.P(z)$$

where the quantifiers are understood to range over A. Saying this another way, if for any element $x \in A$, from the *induction hypothesis* $\forall y.\ y < x \Rightarrow P(y)$ (i. e., the property P holds for all elements less than x) we can prove that $P(x)$ holds, then P holds for every element of A. We can prove that WFI is true for any well founded poset.

Prop 3 *For any well founded poset $\langle A, \leq \rangle$,*

$$\forall x.((\forall y.\ y < x \Rightarrow P(y)) \Rightarrow P(x)) \Rightarrow \forall z.P(z) \qquad \textit{(WFI)}$$

Proof: Assume P is some predicate satisfying the hypothesis of the WFI formula.

$$(1) \quad \forall x.((\forall y.\ y < x \Rightarrow P(y)) \Rightarrow P(x))$$

Let $S \subseteq A$ be the set of elements $a \in A$ such that $\not P(a)$. Assume $S \neq \emptyset$, then since A is well founded, S will have a minimal element s, and by the definition of S that means that $\forall y.\ y < s \Rightarrow P(y)$. But then by (1), it follows that $P(s)$, a contradiction. So $S = \emptyset$, meaning $forallx.P(x)$.□

Example. Let us use this induction principle to prove the following statement about natural numbers.

Prop 4 *If $n \geq 2$ then n is divisible by a prime.*

Proof: Since $\langle \mathbb{N}, \leq \rangle$ is well founded, the Principle of Well Founded Induction applies. Let

$$P(x) \ \equiv \ x < 2 \ \vee \ \exists p \in \text{Primes}.\ p|x$$

We want to show that

$$(1) \quad (\forall y.\ y < x \Rightarrow P(y)) \Rightarrow P(x)$$

So, given an arbitrary $x \in \mathbb{N}$, we assume the *Induction Hypothesis*

$$\text{(IH)} \quad \forall y.\ y < x \Rightarrow P(y)$$

If $x < 2$ then P(x) is immediate. If $2 \leq x$ and x is prime, then x is divisible by a prime, namely itself. If x is composite, then $x = uv$ for some u, v such that $2 \leq u < x$ and $2 \leq v < x$. By (IH), we have P(u), and since $2 \leq u$ we must have $p|u$ for some prime p. But then $p|x$. So we have proved statement (1), and hence by the Principle of Well Founded induction, we can conclude $\forall x.P(x)$. □

This is called a proof by *course-of-values* induction or *complete* induction on the natural numbers. A more common form of induction on the natural numbers is expressed by the following general argument: If we can prove

$$(1) \qquad P(0)$$
$$(2) \qquad \forall n.P(n) \Rightarrow P(n+1)$$

then we can conclude $\forall n.P(x)$. Statement (1) is called the *base* case, and (2) is the *inductive* case. Such a proof is based on the *Principle of Induction over* $\mathbb{N}$ expressed as:

$$P(0) \wedge (\forall n.P(n) \Rightarrow P(n+1)) \Rightarrow \forall n.P(n)$$

Here is an example of a typical proof by induction over $\mathbb{N}$. This should be familiar.

Prop 5 *For any* $n \in \mathbb{N}$*:*

$$\sum_{i=0}^{n} i = \frac{n(n+1)}{2}$$

Proof: We want to prove $\forall n.P(n)$ for the predicate

$$P(n) \equiv \sum_{i=0}^{n} i = \frac{n(n+1)}{2}$$

The argument breaks down into two cases corresponding to the formulas (1) and (2) above.

Base case: n = 0. We show that $P(0)$ by direct calculation:

$$\sum_{i=0}^{0} i = 0 = \frac{0(0+1)}{2}$$

Next we need to prove for an arbitrary number m that $P(m)$ implies $P(m+1)$.

Inductive case: n = m+1. We assume as the *Inductive Hypothesis*, $P(m)$; that is

$$(IH) \quad \sum_{i=0}^{m} i = \frac{m(m+1)}{2}$$

then

$$\begin{aligned}
\sum_{i=0}^{n} i = \left(\sum_{i=0}^{m} i\right) + n &= \frac{m(m+1)}{2} + n \quad \text{(by IH)} \\
&= \frac{m(m+1)}{2} + m + 1 \\
&= \frac{m(m+1)}{2} + \frac{2m+2}{2} \\
&= \frac{m^2+3m+2}{2} \\
&= \frac{(m+1)(m+2)}{2} \\
&= \frac{n(n+1)}{2}
\end{aligned}$$

and we have established the desired $P(n)$.

Now the Induction Principle for $\mathbb{N}$ gives the conclusion $\forall n.P(n)$. □

Next we show that the Principle of Induction over $\mathbb{N}$ is actually a consequence of the more general Principle of Well Founded Induction, given that $\langle \mathbb{N}, \leq \rangle$ is a well founded poset.

Prop 6 *For any unary predicate* P *over* $\mathbb{N}$*,*

$$P(0) \wedge \forall x.(P(x) \Rightarrow P(x+1)) \Rightarrow \forall x.P(x)$$

Proof: We will give a stylized proof as a numbered sequence of justified assertions.

```
(0) $P(0) \ \wedge\ \forall x.(P(x) \Rightarrow P(x+1)) \ \Rightarrow \forall x.P(x)$      TBS
(1) $P(0)$                                                                   Hypo (0)
(2) $\forall x.(P(x) \Rightarrow P(x+1))$                                    Hypo (0)
(3) **Lemma** $\forall x.(\forall y.y < x \Rightarrow P(y)) \Rightarrow P(x)$
    (4) Let $x \in \mathbb{N}$                                               Defn
    (5) $\forall y.y \leq x \Rightarrow P(y)$                                Hypo (3)
    (6) By cases on $n$
        (7) Case $x = 0$
            (8) $P(0)$                                                       (1)
            (9) $P(x)$                                                       (8, Case defn (7))
        (10) Case $x = z + 1$
            (11) $z < x$                                                     (defn <)
            (12) $P(z)$                                                      (5)
            (13) $P(z + 1)$                                                  (2)
            (14) $P(x)$                                                      (Case defn (10))
        (15) $P(x)$                                                          (Cases, 8, 14)
    (16) QED (3)
(17) $\forall x.P(x)$                                                        (3, WFI)
(18) QED (0)                                                                 (17 / 1,2)
```

This streamlined style of presentation of a proof consists of numbered statements constituing the steps of reasoning, with the justification of each step given in the right-hand column. Some lines (e. g., (4)) introduce local variables, some introduce subcase arguments ((7), (10)) and some introduce Lemmas ((3)). Abbreviations are: TBS – "to be shown"; Defn – "defining (or introducing) a local variable; Hypo – "a hypothesis from the current proposition or lemma". The actual inference process for a given line is usually obvious, given the earlier statements from which it is derived. For instance, line (12) is derived by instantiating the universally quantified variable y from line (5) with z and then applying Modus Ponens with the previous line (11). In the common case where a line derives from the previous line, we do not explicitly mention the previous line number in the justification.

3 Terms

Many computational models involve the concept of *terms*[1] In particular, the abstract syntax of programming languages is typically expressed using terms, and computational models often manipulate terms.

Informally, a term is either a constant symbol or the application of a function symbol to a finite sequence of arguments, which are terms. Examples expressed in conventional notation are:

$$A, F(A), F(G(A, B), H(C))$$

For the moment, we are not concerned with assigning *arities* to our function symbols and requiring well-formedness with respect to function arities, so a term like $F(F)$ will be accepted.

More formally, we can define the general universe of terms as follows:

1. $\mathcal{F} = \{A, B, C, \ldots, F, G, H, \ldots\}$ is a (countable) set of function symbols. By convention we use capitalized alphanumeric identifiers for function symbols.

2. $\mathcal{T} = \mathcal{F} \times \mathcal{T}^*$ is the set of general terms, where $\mathcal{T}^*$ denotes the set of all finite sequences of terms (using the Kleene star notation from regular expression).

[1] We also use the word *expressions* as a synonym for terms.

We abbreviate $(A, ())$ as A, calling this an *atomic* term or *constant*, and $(F, (t_1, t_2, \ldots, t_n))$ as $F(t_1, t_2, \ldots, t_n)$, called a *compound* term or function *application*.

What is the meaning (*semantics*) of a definition like this? We can explain it in terms of a fixed-point construction over a universe U, which we assume to be a set containing $\mathcal{F}$ and closed under the formation of finite sequences of elements of U (i.e. if $u_1, \ldots, u_n \in U$, then $(u_1, \ldots, u_n) \in U$.[2]

An alternate way of expressing the inductive definition of general terms is:

1. F is a term, for any $F \in \mathcal{F}$.
2. If $t_1, \ldots, t_n$ are terms and $F \in \mathcal{F}$ then $F(t_1, \ldots, t_n)$ is a term.
3. Nothiing else is a term.

Clause 1 is the base case, clause 2 is the inductive case, and clause 3 expresses *minimality*: $\mathcal{T}$ is the smallest set that is closed with respect to clauses 1 and 2.

We define a *closure function* **Cl** on the powerset of U as follows:

$$\mathbf{Cl} : \mathcal{P}(U) \to \mathcal{P}(U)$$

$$\mathbf{Cl}(X) = \mathcal{F} \times X^* \tag{1}$$

And we note that $\mathbf{Cl}(X)$ is monotonic with respect to the subset ordering on $\mathcal{P}(U)$: $X \subseteq Y \implies \mathbf{Cl}(X) \subseteq \mathbf{Cl}(Y)$. To establish that there is a set X satisfying equation (1), we will invoke a version of the Tarski Fixed Point Theorem.

Defn 6 (least fixed point) *If $\langle A, \leq \rangle$ is a poset, and $F : A \to A$ is a function over A, then $a \in A$ is a* ***fixed point*** *of f if $f(a) = a$. An element a is a* ***least fixed point*** *of f if a is a fixed point of f and if b is also a fixed point of f, then $a \leq b$. A point $x \in A$ is a* ***pre-fixed point*** *of f if $f(x) \leq x$.*

The following is a specialized version of the Tarski (or Tarski-Knaster) theorem proving the existence of least fixed points of a monotonic function on a powerset ordered by the subset relation.

Prop 7 (Tarski) *Given a set S and a monotonic function $f : \mathcal{P}(S) \to \mathcal{P}(S)$ over the poset $\langle \mathcal{P}(S), \subseteq \rangle$, there exists a set $X \in \mathcal{P}(S)$ that is a least fixed point for f.*

Proof: We will define a set P as the intersection of all pre-fixed points of F, and then show that P is actually the least fixed point of F.

(1) F monotonic	Hypo
(2) Let $Pre = \{X \mid F(X) \subseteq X\}$	Defn
(3) Let $P = \bigcap Pre$	Defn
(4) Let $X \in Pre$	Defn
(5) $F(X) \subseteq X$	(pre-fixed point)
(6) $P \subseteq X$	(intersection)
(7) $F(P) \subseteq F(X)$	(1,6,monotonic)
(8) $F(P) \subseteq X$	(5,7,transitivity)
(9) $\forall X \in Pre.\ F(P) \subseteq X$	(8,generalization)
(10) $F(P) \subseteq \bigcap Pre$	(intersection)
(11) $F(P) \subseteq P$	(3)
(12) $F(F(P)) \subseteq F(P)$	(1,11,monotonic)
(13) $F(P) \in Pre$	(2)
(14) $P \subseteq F(P)$	(intersection)

[2] We take it for granted that such a universe U exists, but we can construct such a universe set-theoretically.

(15) $F(P) = P$	(set equality)
(16) P is a fixed-point of F □	
(17) Let Q be a fixed point of F	Defn
(18) $F(Q) = Q$	(17)
(19) $F(Q) \subseteq Q$	(set equality)
(20) $Q \in \mathit{Pre}$	(2)
(21) $P \subseteq Q$	(3,intersection)
(22) P is the least fixed point of F	QED

Now, since we know that **Cl** is monotonic, we can define the set of terms by:

$$\mathcal{T} = \text{least fixed point of } \mathbf{Cl}$$

This notion of term is a bit too general, since it allows all the following to be considered valid terms (where $F \in \mathcal{F}$ is a function symbol):

$$F, \;\; F(F), \;\; F(F,F), \;\; F(F,F,F), \;\; \ldots$$

We normally want a more restricted notion of terms where our function symbols are assigned *arities* (natural numbers) to specify how many arguments they take, and we want terms to be well-formed in the sense that they respect the assigned arities. An assignment of arities is called a signature, and we normally restrict ourselves to a finite set of chosen function symbols. If we want to be explicit, we can indicate the arity of a function symbol as a superscript. So

$$F^2(A^0, B^0), \;\; F^2(A^0, G^1(B^0))$$

are legal, but

$$F^2, \;\; F^2(A^0), \;\; F^2(A^0, G^1, B^0)$$

are not well formed and are not considered valid terms.

3.1 Another Fixed Point Construction

In the proof of Tarski's Theorem, the least fixed point is defined "from the outside" as the intersection of the set of all pre-fixed points of f. Another way of defining the least fixed point is to build it incrementally, "from the inside". We perform the following construction.

We define a sequence of sets P_n as follows:

$$\begin{aligned} P_0 &= \emptyset \\ P_{n+1} &= f(P_n) \end{aligned}$$

Since f is assumed to be monomorphic, we can prove by induction on n that $P_n \subseteq P_{n+1}$. We define P to be the union of the sets P_n:

$$P = \bigcap_{n \in \mathbb{N}} P_n \tag{2}$$

Now assume that

$$f(\bigcap_{n \in \mathbb{N}} P_n) = \bigcap_{n \in \mathbb{N}} f(P_n) \tag{3}$$

then

$$f(P) = f(\bigcap_{n \in \mathbb{N}} P_n) = \bigcap_{n \in \mathbb{N}} f(P_n) = \bigcap_{n \in \mathbb{N}} P_{n+1} = P \tag{4}$$

This is a nice short argument, but the flaw is that equation (4) is not true in all cases. However, it will be true if f is *continous*.

Defn 7 (continous) *A function $f : \mathcal{P}(S) \to \mathcal{P}(S)$ is **continous** if for any increasing sequence of sets $A_0 \subseteq A_1 \subseteq A_2 \ldots$,*

$$f(\bigcap_{n \in \mathbb{N}} A_n) = \bigcap_{n \in \mathbb{N}} f(A_n)$$

So if f is continous as well as monotonic, equation (4) will hold and we

Not all monotonic functions are continous, but, fortunately, there is a useful class of continuous functions that includes most of the closure functions occurring in our inductive definitions.

Defn 8 (finitary function) *A function $f : \mathcal{P}(S) \to \mathcal{P}(S)$ is **finitary** if whenever $a \in f(A)$, there is a finite subset $A_0 \subseteq A$ such that $a \in f(A_0)$.*

Now we can show the following:

Prop 8 *If $f : \mathcal{P}(S) \to \mathcal{P}(S)$ is monotonic and finitary, then it is continuous.*

Proof: Exercise.

So assuming our closure function is finitary (why is this true of **Cl** above?), we have establised that P is a fixed point of f, and it remains to show that P is the least fixed point. So assume Q is some other fixed point of f; we need to show that $P \subseteq Q$:

(0) $P \subseteq Q$	TBS
(1) f monotonic	Hypo
(2) $f(Q) = Q$	Hypo
(3) $\forall n. P_n \subseteq Q$	**Lemma**
(4) by induction on n	
(5) Case n = 0:	case defn
(6) $P_0 = \emptyset \subseteq Q$	defn P_0, sets
(7) Case n = m+1:	case defn
(8) $P_m \subseteq Q$	I.H.
(9) $f(P_m) \subseteq f(Q)$	1, 8
(10) $P_{m+1} \subseteq f(Q)$	defn P_k
(11) $P_n \subseteq f(Q)$	7
(12) $P_n \subseteq Q$	2
(13) QED 3	
(14) $\bigcup\{P_n \mid n \in \mathbb{N}\} \subseteq Q$	3, sets
(15) $P \subseteq Q$	defn P
(16) QED 0	

Since the least fixed point of a function is unique, these two constructions produce the same set. Also if $x \in P$, the least fixed point of f, then we can define $\mathrm{rank}(x)$ to be the least n such that $x \in P_n$.

3.2 Examples

(1) The natural numbers as terms.

Our signature is

$$\mathcal{F} = \{\mathrm{Z}^0, \mathrm{S}^1\}$$

where of course Z^0 represents the constant zero, and S^1 represents the unary successor operation, and we have the correspondence:

$$\begin{array}{lcl} 0 & \rightarrow & \mathrm{Z} \\ 1 & \rightarrow & \mathrm{S(Z)} \\ 2 & \rightarrow & \mathrm{S(S(Z))} \\ & \ldots & \end{array}$$

The set of number terms is defined by the recursive equation

$$\mathcal{N} = \{\mathrm{Z}^0\} \cup (\{\mathrm{S}^1\} \times \mathcal{N})$$

or, $\mathcal{N}$ is the least fixed point of a closure operator

$$\begin{array}{lcl} \mathbf{Cl}(X) & : & \mathcal{U} \rightarrow \mathcal{U} \\ & = & \{\mathrm{Z}^0\} \cup (\{\mathrm{S}^1\} \times X) \end{array}$$

$$\mathcal{N} = \mathbf{Cl}(\mathcal{N})$$

Cl is monotonic because union and product are monotonic. It is also finitary, since a new term is included based on at most one precurser subterm.

We have other standard ways of specifying the set of terms generated inductively over a signature. It is common to use a pseudo-BNF notation like the following:

$$e ::= \mathrm{Z} \mid \mathrm{S}(e)$$

where e is a metavariable ranging over the set of terms being defined. This definition is analagous to the ML datatype declaration

```
datatype e = Z | S of e
```

We can define an ordering on terms in $\mathcal{N}$ derived from the immediate subterm relation, i. e., $t\ R\ \mathrm{S}(t)$. R is an *acyclic binary relation*[3] because if $t_1\ R\ t_2$ we will have $\mathrm{rank}(t_1) < \mathrm{rank}(t_2)$, where rank is defined based on the iterative fixed point construction as discussed above. This also means that we can complete the R relation to obtain a partial ordering $\leq$ on terms by taking reflexive, transitive closureo f R. This partial ordering will also be well founded because an infinite descending chain would imply an infinite decending chain of the corresponding ranks, which are natural numbers, contradicting the fact that $\mathbb{N}$ is well founded. Thus we can prove properties of terms in $\mathcal{N}$ by well founded induction with respect to this partial ordering induced by the immediate subterm relation. This ordering on terms clearly agrees with the standard ordering on the natural numbers corresponding to the terms.

But there is a specialized induction principle for proving that a property P holds for all terms in $\mathcal{N}$:

$$P(\mathrm{Z})\ \wedge\ (\forall t.\, P(t) \Rightarrow P(\mathrm{S}(t)))\ \Rightarrow\ \forall t.\, P(t)$$

As for Proposition 6, this induction principle can be proved using the more general well founded induction principle. The pattern for an inductive proof then will take the following form:

```
Base case: $t = \mathrm{Z}$
    Show $P(\mathrm{Z})$, hence $P(t)$
Ind. case: $t = \mathrm{S}(u)$
    Assume Induction Hypothesis: $P(u)$
    Show $P(\mathrm{S}(u))$, hence $P(t)$
Conclude $\forall t.\, P(t)$
```

[3] Binary relation R is *acyclic* if there is no sequence $a = a_0 R a_1 R a_2\ \ldots\ a_{n-1} R a_n = a$, i. e., there is no cycle in the graph generated by R.