Before starting, be sure that you understand the course policy on Academic Integrity. Download code03.tar from the course website, which contains Standard ML files for a PM interpreter.

1. Determinacy of IMP

Consider the following theorem and proof:

- Determinacy of heap lookup:

(For all \( H, x, c_1, \) and \( c_2 \))

\[ \text{If } H @ x \leadsto c_1 \text{ and } H @ x \leadsto c_2, \text{ then } c_1 = c_2. \]

**Proof:** By structural induction on (the derivation) \( H @ x \leadsto c_1 \).

- \( \rule{H @ x \leadsto c_1 = \text{\textbf{\textsc{empty}}} \cdot @ x \leadsto 0} \)

\[
\begin{align*}
H = \cdot \text{ and } c_1 = 0.
\end{align*}
\]

From \( H @ x \leadsto c_2 \) and \( H = \cdot \), we have \( \cdot @ x \leadsto c_2 \).

By inversion of \( \cdot @ x \leadsto c_2 \), we have \( c_2 = 0 \).

Therefore, \( c_1 = 0 = c_2 \).

- \( \rule{H @ x \leadsto c_1 = \text{\textbf{\textsc{hit}}} \cdot @ x \leadsto c_1 \rightleftharpoons c_1} \)

\[
\begin{align*}
H = H', x \mapsto c_1, c_1 = c_1.
\end{align*}
\]

From \( H @ x \leadsto c_2 \) and \( H = H', x \mapsto c_1 \), we have \( H', x \mapsto c_1 @ x \leadsto c_2 \).

By inversion of \( H', x \mapsto c_1 @ x \leadsto c_2, \) we have \( c_2 = c_1 \).

Therefore, \( c_1 = c_2 \).

- \( \rule{H @ x \leadsto c_1 = \text{\textbf{\textsc{miss}}} H', y \mapsto c' @ x \leadsto c_1} \)

\[
\begin{align*}
H = H', y \mapsto c'.
\end{align*}
\]

From \( H @ x \leadsto c_2 \) and \( H = H', y \mapsto c' \), we have \( H', y \mapsto c' @ x \leadsto c_2 \).

By inversion of \( H', y \mapsto c' @ x \leadsto c_2 \) (with \( x \neq y' \)), we have \( H' @ x \leadsto c_2 \).

By the inductive hypothesis applied \( H' @ x \leadsto c_1 \) with \( H' @ x \leadsto c_2, \) we have \( c_1 = c_2 \).

Therefore, \( c_1 = c_2 \).
(a) Complete the following proof.

- Hint: Use “Determinacy of heap lookup”.

Determinacy of expressions:

(For all $H, e, c_1, $ and $c_2$)

If $H; e \downarrow c_1$ and $H; e \downarrow c_2$, then $c_1 = c_2$.

Proof: By structural induction on (the derivation) $H; e \downarrow c_1$.

- $H; e \downarrow c_1 \equiv H; c_1 \downarrow c_1$ \text{\textsc{const}}

  Therefore, $e = c_1$.

  ***COMPLETE THE \textsc{const} CASE.***

- $H; e \downarrow c_1 \equiv H; e \uparrow x \leadsto c_1$ \text{\textsc{var}}

  Therefore, $e = x$ and $H \uparrow x \leadsto c_1$.

  ***COMPLETE THE \textsc{var} CASE.***

- $H; e \downarrow c_1 \equiv H; e_a \downarrow c_{a1} \quad H; e_b \downarrow c_{b1}$ \text{\textsc{add}}

  Therefore, $e = e_a + e_b$, $c_1 = c_{a1} + c_{b1}$, $H; e_a \downarrow c_{a1}$, and $H; e_b \downarrow c_{b1}$.

  From $H; e \downarrow c_2$ and $e = e_a + e_b$, we have $H; e_a \uparrow e_b \downarrow c_2$.

  By inversion of $H; e_a \uparrow e_b \downarrow c_2$, we have $c_2 = c_{a2} + c_{b2}$, $H; e_a \downarrow c_{a2}$, and $H; e_b \downarrow c_{b2}$.

  (Note: Inverting $H; e_a \uparrow e_b \downarrow c_2$ yielded derivations that evaluate $e_a$ and $e_b$ to constants $c_{a2}$ and $c_{b2}$.)

  ***COMPLETE THE \textsc{add} CASE.***

- $H; e \downarrow c_1 \equiv H; e_a \uparrow e_b \downarrow c_{a1} \uparrow c_{b1}$ \text{\textsc{mult}}

  Therefore, $e = e_a \times e_b$, $c_1 = c_{a1} \times c_{b1}$, $H; e_a \downarrow c_{a1}$, and $H; e_b \downarrow c_{b1}$.

  ***COMPLETE THE \textsc{mult} CASE.***
(b) Complete the following proof.

- Hint: Use “Determinacy of expressions”.

Determinacy of statements:

(For all \( H, s, H'_1, s'_1, H'_2, \) and \( s'_2 \))

If \( H; s \rightarrow H'_1; s'_1 \) and \( H; s \rightarrow H'_2; s'_2 \), then \( H'_1 = H'_2 \) and \( s'_1 = s'_2 \).

Proof: By structural induction on (the derivation) \( H; s \rightarrow H'_1; s'_1 \).

- \( H; s \rightarrow H'_1; s'_1 \) \( \equiv \) \( H; x := e \rightarrow H; x := c_1; \text{skip} \):
  
  Therefore, \( s = x := e, H'_1 = H; x := c_1, s'_1 = \text{skip} \), and \( H; e \downarrow c_1 \).
  
  ***COMPLETE THE ASSIGN CASE.***

- \( H; s \rightarrow H'_1; s'_1 \) \( \equiv \) \( H; \text{while } e \ s_w \rightarrow H; \text{if } e \ (s_w ; \text{while } e \ s_w) \text{ skip} \):
  
  Therefore, \( s = \text{while } e \ s_w, H'_1 = H, \) and \( s'_1 = \text{if } e \ (s_w ; \text{while } e \ s_w) \text{ skip} \).
  
  ***COMPLETE THE WHILE CASE.***

- \( H; s \rightarrow H'_1; s'_1 \) \( \equiv \) \( H; \text{skip} ; s_b \rightarrow H; s_b \):
  
  Therefore, \( s = \text{skip} ; s_b, H'_1 = H, s'_1 = s_b \).
  
  ***COMPLETE THE SEQSKIP CASE.***

- \( H; s_a \rightarrow H'_1; s'_1 \) \( \equiv \) \( H; s_a ; s_b \rightarrow H'_1; s'_1 \):
  
  Therefore, \( s = s_a ; s_b, s'_1 = s'_1 s_a, s_b, \) and \( H; s_a \rightarrow H'_1; s'_1 \).
  
  From \( H; s \rightarrow H'_2; s'_2 \) and \( s = s_a ; s_b, \) we have \( H; s_a ; s_b \rightarrow H'_2; s'_2 \).
  
  By inversion of \( H; s_a ; s_b \rightarrow H'_2; s'_2 \), we have two cases:

  - \( H; s_a ; s_b \rightarrow H'_2; s'_2 \) \( \equiv \) \( H; \text{skip} ; s_b \rightarrow H; s_b \):
    
    Therefore, \( s_a = \text{skip}, H'_2 = H, \) and \( s'_2 = s_b \).
    
    From \( H; s_a \rightarrow H'_1; s'_1 \) and \( s_a = \text{skip}, \) we have \( H; \text{skip} \rightarrow H'_1; s'_1 \).
    
    But, \( H; \text{skip} \rightarrow H'_1; s'_1 \) is impossible.
    
    Therefore, vacuously, \( H'_1 = H'_2 \) and \( s'_1 = s'_2 \).

  - \( H; s_a ; s_b \rightarrow H'_2; s'_2 \) \( \equiv \) \( H; s_a ; s_b \rightarrow H'_2; s'_2 \):
    
    Therefore, \( s'_2 = s'_2, \) and \( H; s_a \rightarrow H'_2; s'_2 \).
    
    ***COMPLETE THE SEQSTEP/SEQSTEP CASE.***

(continued)
• $H; s \rightarrow H'_1; s'_1 \equiv H; \text{if } e \leftarrow s_t \leftarrow s_f \rightarrow H; s_t$:
  Therefore, $s = \text{if } e \leftarrow s_t \leftarrow s_f$, $H'_1 = H$, $s'_1 = s_t$, $H; e \downarrow c_1$, and $c_1 > 0$.

From $H; s \rightarrow H'_2; s'_2$ and $s = \text{if } e \leftarrow s_t \leftarrow s_f$, we have $H; \text{if } e \leftarrow s_t \leftarrow s_f \rightarrow H'_2; s'_2$.

By inversion of $H; \text{if } e \leftarrow s_t \leftarrow s_f \rightarrow H'_2; s'_2$, we have two cases:

- $H; e \downarrow c_2 \quad c_2 > 0$ $wT$
  Therefore, $H'_2 = H$, $s'_2 = s_t$, $H; e \downarrow c_2$, and $c_2 > 0$.
  ***COMPLETE THE ifT/0 CASE.***

- $H; e \downarrow c_2 \quad c_2 \leq 0$ $wF$
  Therefore, $H'_2 = H$, $s'_2 = s_f$, $H; e \downarrow c_2$, and $c_2 \leq 0$.
  ***COMPLETE THE ifF/0 CASE.***

Hint: Use “Determinacy of expressions” to show that this case is impossible.

• $H; s \rightarrow H'_1; s'_1 \equiv H; \text{if } e \leftarrow s_t \leftarrow s_f \rightarrow H; s_f$:
  Therefore, $s = \text{if } e \leftarrow s_t \leftarrow s_f$, $H'_1 = H$, $s'_1 = s_f$, $H; e \downarrow c_1$, and $c_1 \leq 0$.
  ***COMPLETE THE ifF CASE.***
2. PM: A Pattern Matching Language

This problem considers a small “language” for ML-style pattern matching, with some twists. A “program” is a pattern p and a value v. If p “matches” v, then the result is a list of bindings b. If p does not “match” v, then there is no result.

Syntax definition:

\[
\begin{align*}
v &::= c \mid (v, v) \mid s(v) \\
p &::= _ \mid x \mid c \mid (p, p) \mid s(p) \mid \ldots(p) \\
b &::= \cdot \mid b, x \mapsto v \\
(c &\in \{\ldots, -2, -1, 0, 1, 2, \ldots\}) \\
(s &\text{ any non-empty string of letters}) \\
(x &\text{ any non-empty string of letters})
\end{align*}
\]

For values, we have constants, pairs, and tagged values. (The tag is any non-empty string, which is unlike ML where constructors must be declared in preceding \texttt{datatype} declarations.) For patterns, we have wildcard, variables, constants, pairs, tagged patterns, and the “descendent” pattern \ldots(p).

Informal semantics:

- Pattern _ matches every value and produces the empty list of bindings (\cdot).
- Pattern x matches every value and produces the one-element binding list \cdot, x \mapsto v when matched with v. Note x can be any variable.
- Pattern c matches only the value that is the same constant and produces the empty list of bindings.
- Pattern (p_1, p_2) matches only pairs of values and only if p_1 and p_2 match the corresponding components of the pair. The result is the two binding lists from the nested matches appended together.
- Pattern s(p) matches only a tagged value where the tag is the same (i.e., the same string s) and p matches the corresponding value. The result is the result of the nested match.
- Pattern \ldots(p) matches a value v if p matches any descendent of v in the abstract syntax tree, \textit{including v itself}. Put another (very useful) way, it matches if p matches v or if \ldots(p) matches a child of v in the abstract syntax tree. The result is the result of (any) match that leads to “success”.
- Assume a pattern does not have the same variable more than once; you do not need to check for this.

Example: using the concrete syntax for the scanner/parser provided (note parentheses are necessary, the pattern and value must be on separate lines, and there can be no line breaks within the pattern or value):

\[
\begin{align*}
\text{bar (} &\ldots((18, z)), \ldots\text{)} \\
\text{bar (} &42, (\text{foo (}17, (18, (0, 20)))\text{)), 19)\text{)}
\end{align*}
\]

The only match produces a binding list where x maps to 42 and z maps to (0,20).
(a) (Formal large-step operational semantics) Give a formal large-step operational semantics for pattern-matching. Your judgment should have the form \( p;v \Downarrow b \), meaning \( p \) matches \( v \) producing \( b \). If \( p \) does not match \( v \) there must be no derivation for any \( b \). Hints:

- Have 9 inference rules, 3 of which are axioms.
- You need multiple rules for descendent patterns.
- Write \( b_1 @ b_2 \) for the result of appending \( b_1 \) and \( b_2 \).
  (This arises only once; do not worry about formalizing append.)

(b) Give a \( p \) and \( v \) where multiple \( b \) are possible. That is, show the large-step semantics is nondeterministic.

(c) (SML warm-up) Implement \texttt{valueToString} of type \texttt{Ast.value->string} for converting values to concrete syntax. Implement \texttt{bindingListToString} of type \texttt{(string * Ast.value) list -> string} for converting a binding list to a string. The actual string is unimportant; we recommend putting each binding on a separate line and putting a “;” between the variable and the value. Note that the provided \texttt{printAns} uses \texttt{bindingListToString}.

(d) (SML large-step operational semantics) Implement \texttt{large} of type \texttt{Ast.pattern * Ast.value -> (string * Ast.value) list option} to implement pattern-matching. Your code should largely correspond to your inference rules (Hint: use a \texttt{case} match on the pair \( (p,v) \).) with these differences:

- Return \texttt{NONE} if there is no match and \texttt{SOME} \( b \) if there is a match with binding-list \( b \).
- You may resolve the nondeterminism in any manner you like, i.e., if there is more than one match your implementation should just “find one” and return it. Your implementation must always produce one if there is one.

(e) (Formal small-step operational semantics) Give a formal small-step operational semantics for pattern-matching. Your judgment should have the form \( p;v; \cdot \rightarrow^* _-;v';b' \). Pattern-matching is “done” when \( p \) is \(_-\). Otherwise, if \( p \) matches \( v \) there should be a rule that simplifies \( p \) or \( v \) or both by turning them into \( p' \) and \( v' \). The binding list \( b' \) is either \( b \) or something added onto \( b \). A result for the “whole program” \( p \) and \( v \) is a \( b' \) where \( p;v; \cdot \rightarrow^* _-;v';b' \). If \( p \) and \( v \) do not match, there must be no way to derive \( p;v; \cdot \rightarrow^* _-;v';b' \). Hints:

- Have 9 inference rules, 8 of which are axioms.
- For the nonaxiom, “simplify” the “left side” of a pair-match. 1 axiom rule simplifies a pair-match whose left side is the \(_-\) pattern.
- Almost all the axioms produce the same \( b \) they start with.
- A couple axioms will turn a pattern into \(_-\). This is similar to \texttt{IMP’s} assign rule where we turn an assignment into a \texttt{skip}.
- 1 axiom simplifies a tag-match (if the tags match) by just “stripping off the tags”.

Note: These “hints” are perhaps more for “checking your work” than guiding you.

(f) Explain the difficulty that would arise when turning the small-step operational semantics into an ML interpreter, which takes a program consisting of a pattern \( p \) and a value \( v \) and returns a list of bindings \( b \) if and only if \( p;v; \cdot \rightarrow^* _-;v';b \).
Debriefing

- How many hours did you spend on this assignment?
- Would you rate it as easy, moderate, or difficult?
- How deeply do you feel you understand the material it covers (0% – 100%)?
- If you have any other comments about the assignment, then please include them with your submission or send email to mtf@cs.rit.edu.

Submission

All components of the assignment:

- problems 1, 2a, 2b, 2e, 2f, and debriefing in a file named homework03.pdf
- problems 2c and 2d in the file named pm.sml

must be submitted to the Homework 3 Assignment on MyCourses by the due date.

For handwritten components, please use either a document scanner or a mobile scanning app like Adobe Scan or Microsoft Lens.