Chapter Overview

- From $L_1$ to $L_2$: SLD-derivations
- From $L_2$ to Prolog: Backtracking
- The cut in Prolog
From $L_1$ to $L_2$

A program is a set of clauses - no two of which define the same predicate.

A query is a finite sequence of atoms.
WAM$_2$ Memory Layout

Registers

\[
\begin{array}{c}
P \rightarrow \\
CP \rightarrow \\
S \rightarrow \\
H \rightarrow \\
E \rightarrow \\
\end{array}
\]

(low)

Code Area

Heap

Stack

environment

PDL

(high)

Argument and term registers:
$X_1, X_2, \ldots$
The Basic Picture

\[
\begin{align*}
\text{first}(X, [X|\_]) &. \\
\text{second}(X, [\_, X|\_]) &. \\
\text{sequence}(Y,Z) :­ & \text{first}(X,Z), \\
& \text{second}(X,Y).
\end{align*}
\]

\[
\begin{align*}
\text{first}/2: & \quad \text{\texttt{program\_code\_for\_first}} \\
& \quad \text{\texttt{proceed}} \\
\text{second}/2: & \quad \text{\texttt{program\_code\_for\_second}} \\
& \quad \text{\texttt{proceed}} \\
\text{sequence}/2: & \quad \text{\texttt{allocate}} \\
& \quad \text{\texttt{program\_code\_for\_sequence}} \\
& \quad \text{\texttt{query\_code\_for\_first}} \\
& \quad \text{\texttt{call first/2}} \\
& \quad \text{\texttt{query\_code\_for\_second}} \\
& \quad \text{\texttt{call second/2}} \\
& \quad \text{\texttt{deallocate}} \\
\end{align*}
\]

?- \text{sequence}([a,b],[b,c]), \\
\text{sequence}([b,c],[c,d]).

\[
\begin{align*}
\text{allocate} \\
& \quad \text{\texttt{query\_code\_for\_sequence}} \\
& \quad \text{\texttt{call sequence/2}} \\
& \quad \ldots
\end{align*}
\]
Handling Facts

call p/n ≡ CP ← P + instruction_size(P);
        P ← @(p/n);

proceed ≡ P ← CP;

instruction_size(P): number of memory units occupied by the instruction at address P
Handling Clauses

A permanent variable in a clause $H :- B_1, ..., B_m$

:Ú

Y occurs in more than one of these $m$ sets:

$Var(H) \cup Var(B_1), Var(B_2), ..., Var(B_m)$

To emphasize that a variable is permanent, we write $Y_i$.

Permanent variables are cells in the stack area.
The environment frame is used to save local data upon calling a procedure.

<table>
<thead>
<tr>
<th>E</th>
<th>CE</th>
<th>(continuation environment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E + 1</td>
<td>CP</td>
<td>(continuation point)</td>
</tr>
<tr>
<td>E + 2</td>
<td>n</td>
<td>(number of permanent variables)</td>
</tr>
<tr>
<td>E + 3</td>
<td>(Y_1)</td>
<td>(permanent variable 1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E + n + 2</td>
<td>(Y_n)</td>
<td>(permanent variable n)</td>
</tr>
</tbody>
</table>
WAM$_2$ Machine Instructions

\begin{align*}
allocation N & \equiv \quad newE &\leftarrow& E + \text{STACK}[E+2] + 3; \\
& &\quad \text{STACK}[newE] &\leftarrow& E; \\
& &\quad \text{STACK}[newE+1] &\leftarrow& \text{CP}; \\
& &\quad \text{STACK}[newE+2] &\leftarrow& N; \\
& &\quad E &\leftarrow& newE; \\
& &\quad P &\leftarrow& P + \text{instruction\_size}(P); \\
deallocation & \equiv \quad P &\leftarrow& \text{STACK}[E+1]; \\
& &\quad E &\leftarrow& \text{STACK}[E];
\end{align*}
Example

\[ p/2: \text{allocate} \, 2 \quad \% \quad p \]
\[
\begin{align*}
&\text{get\_variable} \, X_3, A_1 \quad \% \quad (X, \\
&\text{get\_variable} \, Y_1, A_2 \quad \% \quad Y) :- \\
&\text{put\_value} \, X_3, A_1 \quad \% \quad q(X, \\
&\text{put\_variable} \, Y_2, A_2 \quad \% \quad Z \\
&\text{call} \, q/2 \quad \% \quad ), \\
&\text{put\_value} \, Y_2, A_1 \quad \% \quad r(Z, \\
&\text{put\_value} \, Y_1, A_2 \quad \% \quad Y \\
&\text{call} \, r/2 \quad \% \quad ) \\
&\text{dallocate} \quad \% \quad .
\end{align*}
\]

\text{WAM}_2 \, \text{Machine Code for } \quad p(X, Y) :- q(X, Z), r(Z, Y).
Handling the Query

The query is encoded like a clause but without the instructions for the missing head.

Prior to calling the code for the query, an initial environment is created:

\[
\begin{align*}
\text{STACK}[E] & \leftarrow E; \\
\text{STACK}[E+1] & \leftarrow \text{CP}; \\
\text{STACK}[E+2] & \leftarrow 0;
\end{align*}
\]

Following the (successful) call to the query, the register bindings are printed as the CAS.
From $L_2$ to Pure Prolog

A program is an arbitrary set of clauses.
Choice Points vs. Environment Frames

Example:

\begin{verbatim}
a :- b(X), c(X).
b(X) :- e(X).
c(1).
e(X) :- f(X).
e(X) :- g(X).
f(2).
g(1).
?- a.
\end{verbatim}

This example shows that a choice point must protect each previously stored environment frame from deallocation in order that the environment is still available upon backtracking.
WAM Memory Layout

Registers

- P → Code Area (low)
- CP → Heap
- S, HB → Stack
- H → choice point
- B → environment
- E → Trail
- TR → PDL (high)

Argument and term registers:
- $X_1, X_2, \ldots$

- n arity
- A1 1st argument
- \cdot
- An nth argument
- CE cont. environment
- CP cont. point
- B previous choice pt.
- BP next clause
- TR trail pointer
- H heap pointer
Trail

Backtracking requires to “undo” substitutions made after the choicepoint to which we jump back.

The trail contains the addresses of all variables which need to be unbound upon backtracking.

Register HB contains the value of H at the time of the latest choice point. Only variables below HB need to be reset.
Recording the Bindings

procedure bind\(a_1, a_2: address\);

begin

    case \(\text{STORE}[a_1]\) of

    \(<\text{REF},_\>\) : \(\text{STORE}[a_1] \leftarrow \text{STORE}[a_2]\);
        trail\(a_1\);

    other : \(\text{STORE}[a_2] \leftarrow \text{STORE}[a_1]\);
        trail\(a_2\)

    endcase

end bind
Trail Functions

- trail \((addr: address)\)

pushes \(addr\) onto the trail, provided that

\[ addr < HB \]

- unwind_trail \((addr_1, addr_2: address)\)

resets all variables stored in \(addr_1, addr_1 + 1, ..., addr_2 - 1\)
procedure trail(a: address);
    begin
        if (a < HB)
            then
                begin
                    TRAIL[TR] ← a;
                    TR ← TR+1
                end
    end trail

procedure unwind_trail(a_1,a_2: address);
    begin
        for i ← a_1 to a_2 – 1 do
            STORE[TRAIL[i]] ← (REF,TRAIL[i])
    end unwind_trail
The choice point frame is used to save local data needed upon backtracking.

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>n</td>
<td><em>(number of arguments)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B+1</td>
<td>A1</td>
<td><em>(argument register 1)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B+n</td>
<td>An</td>
<td><em>(argument register n)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B+n+1</td>
<td>CE</td>
<td><em>(continuation environment)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B+n+2</td>
<td>CP</td>
<td><em>(continuation point)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B+n+3</td>
<td>B</td>
<td><em>(previous choice point)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B+n+4</td>
<td>BP</td>
<td><em>(next clause)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B+n+5</td>
<td>TR</td>
<td><em>(trail pointer)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B+n+6</td>
<td>H</td>
<td><em>(heap pointer)</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
allocate N ≡ if E > B

  then newE ← E+STACK[E+2] + 3

  else newE ← B+STACK[B] + 7;

  STACK[newE] ← E;

  STACK[newE+1] ← CP;

  STACK[newE+2] ← N;

  E ← newE;

  P ← P + instruction_size(P);
Coding Scheme for Backtracking

\[
\begin{align*}
p(t_{11}, \ldots, t_{1n}) & : B_1 \\
p(t_{21}, \ldots, t_{2n}) & : B_2 \\
p(t_{31}, \ldots, t_{3n}) & : B_3 \\
& \ldots \\
p(t_{k1}, \ldots, t_{kn}) & : B_k
\end{align*}
\]

\[
\begin{align*}
p / n & : \text{try}_\text{me}_\text{else} L_1 \\
& \quad \text{code for first clause} \\
L_1 & : \text{retry}_\text{me}_\text{else} L_2 \\
& \quad \text{code for second clause} \\
& \ldots \\
L_{k-2} & : \text{retry}_\text{me}_\text{else} L_{k-1} \\
& \quad \text{code for third clause} \\
L_{k-1} & : \text{trust}_\text{me} \\
& \quad \text{code for k-th clause}
\end{align*}
\]
try_me_else L ≡ if E > B
    then newB ← E+STACK[E+2] + 3
    else newB ← B+STACK[B] + 7;
STACK[newB] ← num_of_args;
n ← STACK[newB];
for i ← 1 to n do STACK[newB+i] ← Ai;
STACK[newB+n+1] ← E;
STACK[newB+n+2] ← CP;
STACK[newB+n+3] ← B;
STACK[newB+n+4] ← L;
STACK[newB+n+5] ← TR;
STACK[newB+n+6] ← H;
B ← newB;
HB ← H;
P ← P + instruction_size(P);
WAM Machine Instructions (III)

\[
\text{retry\_me\_else } L \equiv \ n \leftarrow \text{STACK}[B]; \\
\text{for } i \leftarrow 1 \text{ to } n \text{ do } A_i \leftarrow \text{STACK}[B+i]; \\
E \leftarrow \text{STACK}[B+n+1]; \\
CP \leftarrow \text{STACK}[B+n+2]; \\
\text{STACK}[B+n+4] \leftarrow L; \\
\text{unwind\_trail}(\text{STACK}[B+n+5], TR); \\
TR \leftarrow \text{STACK}[B+n+5]; \\
H \leftarrow \text{STACK}[B+n+6]; \\
HB \leftarrow H; \\
P \leftarrow P + \text{instruction\_size}(P);
\]
trust_me ≡ 

\[ n \leftarrow \text{STACK}[B]; \]

\[ \text{for } i \leftarrow 1 \text{ to } n \text{ do } A_i \leftarrow \text{STACK}[B+i]; \]

\[ E \leftarrow \text{STACK}[B+n+1]; \]

\[ \text{CP} \leftarrow \text{STACK}[B+n+2]; \]

\[ \text{unwind\_trail(STACK}[B+n+5], TR); \]

\[ \text{TR} \leftarrow \text{STACK}[B+n+5]; \]

\[ H \leftarrow \text{STACK}[B+n+6]; \]

\[ B \leftarrow \text{STACK}[B+n+3]; \]

\[ \text{HB} \leftarrow \text{STACK}[B+n+6]; \]

\[ P \leftarrow P + \text{instruction\_size}(P); \]
backtrack ≡ P ← STACK[B + STACK[B]+4];
(whenever failure occurs)

call p/n ≡ CP ← P + instruction_size(P);
num_of_arg ← n;
P ← @ (p/n);
The Cut

The effect of the cut “!” is to forget any other alternative for the clause in which it appears and any other alternative arising from preceding body atoms in this clause.

This means to discard all choice points (CPs) created after the CP that was the current CP at the time the clause with the cut was called.

An additional cut register $B_0$ is used to store the appropriate CP where to return upon backtracking over a cut.
Call Revisited

call \ p/n \ \equiv \ CP \leftarrow P + instruction\_size(P);\\
num\_of\_arg \leftarrow n;\\
B0 \leftarrow B;\\
P \leftarrow @ (p/n);

The value of \texttt{BO} is saved in every choice point frame.
The Cut as First Body Atom

\[ a/0: \] allocate 0 % \( a \)
\[ \text{neck\_cut} \] % \(-!\),
\[ \text{call} \ b/0 \] % \( b \)
\[ \text{deallocate} \] % .

WAM Machine Code for \( a :- !, b \).
Machine Instruction

\[
\text{neck_cut} \equiv \text{if } B > B0 \text{ then begin}
\begin{align*}
B &\leftarrow B0; \\
HB &\leftarrow \text{STACK}[B+\text{STACK}[B]+6]; \\
tidy\_trail &\end{align*}
\text{end;}
\begin{align*}
P &\leftarrow P + \text{instruction\_size}(P);
\end{align*}
\]

The value of $B0$ is saved in every choice point frame.
procedure tidy_trail;
    begin
        $i \leftarrow \text{STACK}[B + \text{STACK}[B]+5]$;
        while $i < \text{TR}$ do
            if TRAIL[$i$] $<$ HB then
                $i \leftarrow i+1$
            else begin
                TRAIL[$i$] $\leftarrow$ TRAIL[TR-1];
                TR $\leftarrow$ TR-1;
            end
        end
    end tidy_trail;
A Deep Cut

\( a/0: \) allocate 1 % \( a \)
get_level \( Y_1 \)
call \( b/0 \) % \( \vdash b, \)
cut \( Y_1 \) % \( !, \)
call \( c/0 \) % \( c \)
deallocate % .

WAM Machine Code for \( a \vdash b, !, c. \)
Machine Instructions

\[ \text{get\_level } Y_n \equiv \text{STACK}[E+2+n] \leftarrow B0; \]
\[ P \leftarrow P + \text{instruction\_size}(P); \]

\[ \text{cut } Y_n \equiv \text{if } B > \text{STACK}[E+2+n] \text{ then begin} \]
\[ B \leftarrow \text{STACK}[E+2+n]; \]
\[ \text{HB } \leftarrow \text{STACK}[B+\text{STACK}[B]+6]; \]
\[ \text{tidy\_trail} \]
\[ \text{end}; \]
\[ P \leftarrow P + \text{instruction\_size}(P); \]
Objectives

- From $L_1$ to $L_2$: SLD-derivations
- From $L_2$ to Prolog: Backtracking
- The cut in Prolog