CS420
Reducibility
Tuesday November 15, 2022

```
DEFINE DOESITHALT(PROGRAM):
{
    RETURN TRUE;
}
```

THE BIG PICTURE SOLUTION TO THE HALTING PROBLEM
Announcements

• HW 8 in
  • Due Mon 11/14 11:59pm

• HW 9 out
  • Due Mon 11/21 11:59pm
Last Time: Undecidability Proofs

• We proved $\mathcal{A}_{TM} = \{\langle M, w \rangle | M \text{ is a TM and } M \text{ accepts } w \}$ undecidable ...

• ... by contradiction:
  • Use hypothetical $\mathcal{A}_{TM}$ decider to create an impossible decider “D”!

• Step # 1: coming up with “D” --- hard!
  • Need to invent diagonalization

• Step # 2: “reduce” $\mathcal{A}_{TM}$ to the “D” problem --- easier!

• From now on: undecidability proofs only need step # 2!
  • And we now have two “impossible” problems to choose from
Last Time: The Halting Problem

\[ \text{HALT}_{TM} = \{ \langle M, w \rangle | M \text{ is a TM and } M \text{ halts on input } w \} \]

**Thm:** \( \text{HALT}_{TM} \) is undecidable

**Proof**, by contradiction:

- **Assume:** \( \text{HALT}_{TM} \) has decider \( R \); use it to create decider for \( A_{TM} \):
  
  \[ A_{TM} = \{ \langle M, w \rangle | M \text{ is a TM and } M \text{ accepts } w \} \]

- ...

- But \( A_{TM} \) is undecidable and has no decider!
Last Time: The Halting Problem

**Thm:** $HALT_{TM}$ is undecidable

**Proof**, by contradiction:

- **Assume:** $HALT_{TM}$ has *decider* $R$; use it to create decider for $A_{TM}$:

  $$A_{TM} = \{\langle M, w \rangle | M \text{ is a TM and } M \text{ accepts } w\}$$

  $S$ = “On input $\langle M, w \rangle$, an encoding of a TM $M$ and a string $w$:
  1. Run TM $R$ on input $\langle M, w \rangle$.
  2. If $R$ rejects, *reject*.
  3. If $R$ accepts, simulate $M$ on $w$ until it halts.
  4. If $M$ has accepted, *accept*; if $M$ has rejected, *reject*.”

**Termination argument:**
- **Step 1:** $R$ is a decider so always halts
- **Step 3:** $M$ always halts because $R$ said so

Using our hypothetical $HALT_{TM}$ decider $R$
Thm: $\text{HALT}_\text{TM}$ is undecidable

Proof, by contradiction:

• **Assume**: $\text{HALT}_\text{TM}$ has **decider** $R$; use it to create decider for $A_{\text{TM}}$:

  $$A_{\text{TM}} = \{\langle M, w \rangle | M \text{ is a } \text{TM and } M \text{ accepts } w\}$$

  $S =$ “On input $\langle M, w \rangle$, an encoding of a $\text{TM } M$ and a string $w$:

  1. Run $\text{IM } R$ on input $\langle M, w \rangle$.
  2. If $R$ rejects, reject.
  3. If $R$ accepts, simulate $M$ on $w$ until it halts.
  4. If $M$ has accepted, accept; if $M$ has rejected, reject.”

• But $A_{\text{TM}}$ is undecidable! i.e., this decider does not exist!

  • So $\text{HALT}_\text{TM}$ is also undecidable!
Interlude: Reducing from $HALT_{TM}$

A practical thought experiment ... about compiler optimizations

Your compiler changes your program!

If TRUE then A else B $\rightarrow$ A

$1 + 2 + 3 \rightarrow 6$
Compiler Optimizations

Optimization - docs

- **-00**
  - No optimization, faster compilation time, better for debugging builds.
- **-02**
- **-03**
  - Higher level of optimization. Slower compile-time, better for production builds.
- **-O Fast**
  - Enables higher level of optimization than (-03). It enables lots of flags as can be seen in src (-ffloat-store, -ffast-math, -ffinite-math-only, -O3 ...)
- **-finline-functions**
- **-m64**
- **-funroll-loops**
- **-fvecto**
- **-fprofile-generate**

Types of optimization

Techniques used in optimization can be broken up among various scopes which can affect anything from a single statement to the entire program. Generally speaking, locally scoped techniques are easier to implement than global ones but result in smaller gains. Some examples of scopes include:

**Peephole optimizations**

These are usually performed late in the compilation process after machine code has been generated. This form of optimization examines a few adjacent instructions (like “looking through a peephole” at the code) to see whether they can be replaced by a single instruction or a shorter sequence of instructions.[1] For instance, a multiplication of a value by 2 might be more efficiently executed by left-shifting the value or by adding the value to itself (this example is also an instance of strength reduction).

**Local optimizations**

These only consider information local to a basic block.[2] Since basic blocks have no control flow, these optimizations need very little analysis, saving time and reducing storage requirements. However, this also means that no information is preserved across jumps.

**Global optimizations**

These are also called “interprocedural methods” and act on whole functions.[3] This is not just function information about the function but often makes expensive computations necessary. Worst case assumptions have to be made when function calls occur or global variables are accessed because little information about them is available.

**Loop optimizations**

These act on the statements which make up a loop, such as a for loop, for example loop-invariant code motion. Loop optimizations can have a significant impact because many programs spend a large percentage of their time inside loops.[4]

**Prescient store optimizations**

These allow store operations to occur earlier than would otherwise be permitted in the context of threads and locks. The process needs some way of knowing ahead of time what value will be stored by the assignment that it should have followed. The purpose of this relaxation is to allow compiler optimization to perform certain kinds of code rearrangement that preserve the semantics of properly synchronized programs.[5]

**Interprocedural, whole-program or link-time optimization**

These analyze all of a program’s source code. The greater quantity of information extracted means that optimizations can be more effective compared to when they only have access to local information, i.e., within a single function. This kind of optimization can also allow new techniques to be performed. For instance, function inlining, where a call to a function is replaced by a copy of the function body.

**Machine code optimization and object code optimizer**

These analyze the executable task image of the program after all of an executable machine code has been linked. Some of the techniques that can be applied in a more limited scope, such as macro expansion which saves space by collapsing common sequences of instructions, are more effective when the entire executable task image is available for analysis.
**The Optimal Optimizing Compiler**

**Thm**: The Optimal (C++) Optimizing Compiler does not exist

**Proof**, by contradiction:

**Assume**: $OPT$ is the Perfect Optimizing Compiler

Use it to create $HALT_{TM}$ decider (accepts $<M, w>$ if $M$ halts with $w$, else reject):

$S = \text{On input } <M, w>$, where $M$ is C++ program and $w$ is string:

- If $OPT(M) == \text{for}(;;)$
  - a) Then **Reject**
  - b) Else **Accept**
Summary: The Limits of Algorithms

- $A_{DFA} = \{\langle B, w \rangle | B \text{ is a DFA that accepts input string } w\}$: Decidable
- $A_{CFG} = \{\langle G, w \rangle | G \text{ is a CFG that generates string } w\}$: Decidable
- $A_{TM} = \{\langle M, w \rangle | M \text{ is a TM and } M \text{ accepts } w\}$: Undecidable
- $HALT_{TM} = \{\langle M, w \rangle | M \text{ is a TM and } M \text{ halts on input } w\}$: Undecidable
- $E_{DFA} = \{\langle A \rangle | A \text{ is a DFA and } L(A) = \emptyset\}$: Decidable
- $E_{CFG} = \{\langle G \rangle | G \text{ is a CFG and } L(G) = \emptyset\}$: Decidable
- $E_{TM} = \{\langle M \rangle | M \text{ is a TM and } L(M) = \emptyset\}$: Undecidable
Reducibility: Modifying the TM

**Thm:** \( E_{\text{TM}} \) is undecidable

**Proof**, by contradiction:

- Assume \( E_{\text{TM}} \) has decider \( R \); use it to create decider for \( A_{\text{TM}} \):

  \[
  S = \text{“On input } \langle M, w \rangle, \text{ an encoding of a TM } M \text{ and a string } w:\n  \]
  
  - First, construct \( M_1 \)
  
  - Run \( R \) on input \( \langle M_1 \rangle \)
  
  - If \( R \) accepts, reject (because it means \( \langle M \rangle \) doesn’t accept \( w \))
  
  - If \( R \) rejects, then accept (\( \langle M \rangle \) accepts \( w \))

- **Idea:** Wrap \( \langle M \rangle \) in a new TM that can only accept \( w \):

\[
M_1 = \text{“On input } x:\n  
  1. If } x \neq w, \text{ reject.} 
  
  2. If } x = w, \text{ run } M \text{ on input } w \text{ and accept if } M \text{ does.”}
\]

\( M_1 \) accepts \( w \) if \( M \) does.
Reducibility: Modifying the TM

Thm: \( E_{\text{TM}} \) is undecidable

Proof, by contradiction:

• Assume \( E_{\text{TM}} \) has decider \( R \); use it to create decider for \( A_{\text{TM}} \):

  \[
  S = \text{"On input } \langle M, w \rangle \text{, an encoding of a TM } M \text{ and a string } w:\n  \]
  
  1. Run \( R \) on input \( \langle M \rangle \)
  2. If \( R \) accepts, reject (because it means \( \langle M \rangle \) doesn’t accept \( w \))
  3. If \( R \) rejects, then \( \langle M \rangle \) accepts \( \langle M \rangle \) accepts \( w \)

• Idea: Wrap \( \langle M \rangle \) in a new TM that can only accept \( w \):
Summary: The Limits of Algorithms

- $A_{DFA} = \{ \langle B, w \rangle | B$ is a DFA that accepts input string $w \}$
- $A_{CFG} = \{ \langle G, w \rangle | G$ is a CFG that generates string $w \}$
- $A_{TM} = \{ \langle M, w \rangle | M$ is a TM and $M$ accepts $w \}$
- $E_{DFA} = \{ \langle A \rangle | A$ is a DFA and $L(A) = \emptyset \}$
- $E_{CFG} = \{ \langle G \rangle | G$ is a CFG and $L(G) = \emptyset \}$
- $E_{TM} = \{ \langle M \rangle | M$ is a TM and $L(M) = \emptyset \}$
- $EQ_{DFA} = \{ \langle A, B \rangle | A$ and $B$ are DFAs and $L(A) = L(B) \}$
- $EQ_{CFG} = \{ \langle G, H \rangle | G$ and $H$ are CFGs and $L(G) = L(H) \}$
- $EQ_{TM} = \{ \langle M_1, M_2 \rangle | M_1$ and $M_2$ are TMs and $L(M_1) = L(M_2) \}$

Decidable
Decidable
Undecidable
Decidable
Decidable
Undecidable
Decidable
Undecidable
Undecidable

next
Reduce to something else: $EQ_{TM}$ is undecidable

$EQ_{TM} = \{\langle M_1, M_2 \rangle \mid M_1 \text{ and } M_2 \text{ are TMs and } L(M_1) = L(M_2)\}$

Proof, by contradiction:

• Assume: $EQ_{TM}$ has decider $R$; use it to create decider for $A_{TM}$:

$E_{TM} = \{\langle M \rangle \mid M \text{ is a TM and } L(M) = \emptyset\}$

$S =$ “On input $\langle M \rangle$, where $M$ is a TM:

1. Run $R$ on input $\langle M, M_1 \rangle$, where $M_1$ is a TM that rejects all inputs.
2. If $R$ accepts, accept; if $R$ rejects, reject.”
Reduce to something else: $EQ_{TM}$ is undecidable

\[ EQ_{TM} = \{ \langle M_1, M_2 \rangle | M_1 \text{ and } M_2 \text{ are TMs and } L(M_1) = L(M_2) \} \]

Proof, by contradiction:

• **Assume:** $EQ_{TM}$ has decider $R$; use it to create decider for $E_{TM}$:

\[ = \{ \langle M \rangle | M \text{ is a TM and } L(M) = \emptyset \} \]

\[ S = \text{“On input } \langle M \rangle \text{, where } M \text{ is a TM:} \]

1. Run $R$ on input $\langle M, M_1 \rangle$, where $M_1$ is a TM that rejects all inputs.
2. If $R$ accepts, accept; if $R$ rejects, reject.”

• But $E_{TM}$ is undecidable!
Summary: Undecidability Proof Techniques

• **Proof Technique #1:**
  - Use hypothetical decider to implement impossible $A_{TM}$ decider
  - Example Proof: $A_{TM} = \{ \langle M, w \rangle | M \text{ is a TM and } M \text{ accepts } w \}$

• **Proof Technique #2:**
  - Use hypothetical decider to implement impossible $A_{TM}$ decider
  - But first modify the input $M$
  - Example Proof: $E_{TM} = \{ \langle M \rangle | M \text{ is a TM and } L(M) = \emptyset \}$

• **Proof Technique #3:**
  - Use hypothetical decider to implement non-$A_{TM}$ impossible decider
  - Example Proof: $EQ_{TM} = \{ \langle M_1, M_2 \rangle | M_1 \text{ and } M_2 \text{ are TMs and } L(M_1) = L(M_2) \}$
Summary: Decidability and Undecidability

- \( A_{DFA} = \{\langle B, w \rangle \mid B \text{ is a DFA that accepts input string } w \} \) Decidable
- \( A_{CFG} = \{\langle G, w \rangle \mid G \text{ is a CFG that generates string } w \} \) Decidable
- \( A_{TM} = \{\langle M, w \rangle \mid M \text{ is a TM and } M \text{ accepts } w \} \) Undecidable
- \( E_{DFA} = \{\langle A \rangle \mid A \text{ is a DFA and } L(A) = \emptyset \} \) Decidable
- \( E_{CFG} = \{\langle G \rangle \mid G \text{ is a CFG and } L(G) = \emptyset \} \) Decidable
- \( E_{TM} = \{\langle M \rangle \mid M \text{ is a TM and } L(M) = \emptyset \} \) Undecidable
- \( EQ_{DFA} = \{\langle A, B \rangle \mid A \text{ and } B \text{ are DFAs and } L(A) = L(B) \} \) Decidable
- \( EQ_{CFG} = \{\langle G, H \rangle \mid G \text{ and } H \text{ are CFGs and } L(G) = L(H) \} \) Undecidable
- \( EQ_{TM} = \{\langle M_1, M_2 \rangle \mid M_1 \text{ and } M_2 \text{ are TMs and } L(M_1) = L(M_2) \} \) Undecidable
Also Undecidable ...

- $\text{REGULAR}_{TM} = \{<M> \mid M \text{ is a TM and } L(M) \text{ is a regular language}\}$
Thm: $REGULAR_{TM}$ is undecidable

$REGULAR_{TM} = \{ \langle M \rangle | M$ is a TM and $L(M)$ is a regular language $\}$

Proof, by contradiction:

• **Assume:** $REGULAR_{TM}$ has decider $R$; use it to create decider for $A_{TM}$:

  $S = \text{“On input } \langle M, w \rangle, \text{ an encoding of a TM } M \text{ and a string } w:\$
  
  • First, construct $M_2(??)$
  • Run $R$ on input $\langle M_2 \rangle$
  • If $R$ accepts, accept; if $R$ rejects, reject

Want: $L(M_2) =$

• **regular**, if $M$ accepts $w$
• **nonregular**, if $M$ does not accept $w$
Thm: $\text{REGULAR}_{TM}$ is undecidable (continued)

$\text{REGULAR}_{TM} = \{\langle M \rangle | M$ is a TM and $L(M)$ is a regular language $\}$

$M_2 =$ “On input $x$:
1. If $x$ has the form $0^n1^n$, accept.
2. If $x$ does not have this form, run $M$ on input $w$ and accept if $M$ accepts $w$.”

Want: $L(M_2) =$
- **regular**, if $M$ accepts $w$
- **nonregular**, if $M$ does not accept $w$

Always accept strings $0^n1^n$
$L(M_2) =$ **nonregular**, so far

If $M$ accepts $w$, accept everything else, so $L(M_2) = \Sigma^* =$ **regular**

if $M$ does not accept $w$, $M_2$ accepts all strings (**regular** lang)

All strings

$0^n1^n$

if $M$ accepts $w$, $M_2$ accepts this **nonregular** lang
Also Undecidable ...

- \( \text{REGULAR}_{\text{TM}} = \{<M> | M \text{ is a TM and } L(M) \text{ is a regular language}\} \)

- \( \text{CONTEXTFREE}_{\text{TM}} = \{<M> | M \text{ is a TM and } L(M) \text{ is a CFL}\} \)

- \( \text{DECIDABLE}_{\text{TM}} = \{<M> | M \text{ is a TM and } L(M) \text{ is a decidable language}\} \)

- \( \text{FINITE}_{\text{TM}} = \{<M> | M \text{ is a TM and } L(M) \text{ is a finite language}\} \)

Seems like no algorithm can compute anything about language of TMs, i.e., about programs ...
An Algorithm About Program Behavior?

Write a program that, given another program as its argument, returns TRUE if that argument prints “Hello, World!”

```
main()
{
    printf("hello, world\n");
}
```

TRUE
Write a program that, given another program as its argument, returns \texttt{TRUE} if that argument prints \texttt{"Hello, World!"}
Also Undecidable ...

- $\text{REGULAR}_{\text{TM}} = \{ <M> \mid M \text{ is a TM and } L(M) \text{ is a regular language} \}$

- $\text{CONTEXTFREE}_{\text{TM}} = \{ <M> \mid M \text{ is a TM and } L(M) \text{ is a CFL} \}$

- $\text{DECIDABLE}_{\text{TM}} = \{ <M> \mid M \text{ is a TM and } L(M) \text{ is a decidable language} \}$

- $\text{FINITE}_{\text{TM}} = \{ <M> \mid M \text{ is a TM and } L(M) \text{ is a finite language} \}$

- ...

- $\text{ANYTHING}_{\text{TM}} = \{ <M> \mid M \text{ is a TM and “… anything …” about } L(M) \}$

Seems like no algorithm can compute anything about Turing Machines, i.e., about programs ...

Rice’s Theorem
Rice’s Theorem: \( \text{ANYTHING}_{\text{TM}} \) is Undecidable

\[ \text{ANYTHING}_{\text{TM}} = \{<M> \mid M \text{ is a TM and … anything … about } L(M) \} \]

• “... Anything ...”, more precisely:
  • For any \( M_1, M_2 \), if \( L(M_1) = L(M_2) \) ...
  • ... then \( M_1 \in \text{ANYTHING}_{\text{TM}} \iff M_2 \in \text{ANYTHING}_{\text{TM}} \)

• Also, “... Anything ...” must be “non-trivial”:
  • \( \text{ANYTHING}_{\text{TM}} \neq \{\} \)
  • \( \text{ANYTHING}_{\text{TM}} \neq \text{set of all TMs} \)
Rice’s Theorem: $\text{ANYTHING}_{TM}$ is Undecidable

$\text{ANYTHING}_{TM} = \{<M> \mid M \text{ is a TM and … anything … about } L(M)\}$

Proof by contradiction

• **Assume** some language satisfying $\text{ANYTHING}_{TM}$ has a decider $R$.
  • Since $\text{ANYTHING}_{TM}$ is non-trivial, then there exists $M_{\text{ANY}} \in \text{ANYTHING}_{TM}$
  • Where $R$ accepts $M_{\text{ANY}}$

• Use $R$ to create decider for $A_{TM}$:

  **On input $<M, w>$:**
  
  • **Create** $M_w$:
    
    $M_w = \text{on input } x:
    - \text{Run } M \text{ on } w
    - \text{If } M \text{ rejects } w: \text{ reject } x
    - \text{If } M \text{ accepts } w:
      \text{ Run } M_{\text{ANY}} \text{ on } x \text{ and accept if it accepts, else reject}
  
  • **Run** $R$ on $M_w$
    
    • If it accepts, then $M_w = M_{\text{ANY}}$, so $M$ accepts $w$, so accept
    • Else reject

  **Wait!** What if the TM that accepts nothing is in $\text{ANYTHING}_{TM}$?!

  These two cases must be different, (so $R$ can distinguish when $M$ accepts $w$)

  Proof still works! Just use the complement of $\text{ANYTHING}_{TM}$ instead!
Rice’s Theorem Implication

\[ \{<M> \mid M \text{ is a TM that installs malware}\} \]

Undecidable!
(by Rice’s Theorem)
\[ A_{\text{DFA}} = \{ \langle B, w \rangle | B \text{ is a DFA that accepts input string } w \} \]
\[ A_{\text{CFG}} = \{ \langle G, w \rangle | G \text{ is a CFG that generates string } w \} \]
\[ A_{\text{TM}} = \{ \langle M, w \rangle | M \text{ is a TM and } M \text{ accepts } w \} \]

Decidable
Decidable
Undecidable

• In hindsight, of course a restricted TM (a \textit{decider}) shouldn’t be able to simulate unrestricted TM (a \textit{recognizer})

• But could a restricted TM simulate an even more restricted TM?
  • Next time
Check-in Quiz 11/15

On gradescope