CPSC 351 — Assignment #2

Turing Machines, Recognizability, and Decidability

This assignment is to be completed by groups of either three or four students. Groups will be created by the instructor, and each student will be notified of their group, and its other members, on or before ***Wednesday, October 25***.

Please read and make sure that you understand the following material ***before*** you start work on this assignment.

- Information about ***effective teamwork*** on the "Assignments" page of the course web site.
- Information about ***academic integrity*** on the "Assignments" page of the course web site. This includes "Assignment Do's and Don'ts": Information about what is allowed — and what is not allowed — when working on assignments in this course.
- Expectations concerning the ***quality of work*** that you submit for assessment in this course. This is also available on the "Assignments" page of the course web site.

Problem To Be Solved

Consider the alphabet Σ_{TM} that was used for encodings of Turing machines, along with Turing machines and input strings for them, in Lecture #12. Let $\Sigma_{\text{2TM}} = \Sigma_{\text{TM}} \cup \{\#\}$. A Turing machine

$$M = (Q, \Sigma, \Gamma, \delta, q_0, q_{\text{accept}}, q_{\text{reject}})$$

and a state $q \in Q$ can be encoded by the string

$$\mu\#\nu \in \Sigma_{\text{TM}}^{\star}$$

where $\mu \in \Sigma_{\text{TM}}^{\star}$ is the encoding of the Turing machine M, and ν is the encoding of the state $q \in Q$, as these are described in Lecture #12. Now let TM+State $\subseteq \Sigma_{\text{2TM}}^{\star}$ be the set of all encodings of Turing machines and states.

- Since every string $\mu \in$ TM+State includes exactly one copy of “#” — which separates an encoding of a Turing machine M and a state in this Turing machine — and since the language TM $\subseteq \Sigma_{\text{TM}}^\star$ (of encodings of Turing machines) is decidable, it is possible to use the information about encodings, in Lecture #12, to show that the language TM+State is also decidable.

Once again, suppose that

$$M = (Q, \Sigma, \Gamma, \delta, q_0, q_{\text{accept}}, q_{\text{reject}})$$

is a Turing machine. A state $q \in Q$ is ***useful*** if there exists at least one input string $\omega \in \Sigma^\star$ such that the state q is reached during the execution of M on ω — this is, if

$$q_0\omega \vdash^\star \mu_1 q \mu_2$$

for some pair of strings $\mu_1, \mu_2 \in \Gamma^\star$. A state $q \in Q$ is ***useless*** if it is not useful. In other words, q is ***useless*** if it is *not* true that

$$q_0\omega \vdash^\star \mu_1 q \mu_2$$

for *any* pair of strings $\mu_1, \mu_2 \in \Gamma^\star$ and for *any* input string $\omega \in \Sigma^\star$.

Now let UselessState$_{\text{TM}} \subseteq \Sigma_{\text{2TM}}^\star$ be the set of encodings of Turing machines M and states q, of M, such that q is useless. Then

$$\text{UselessState}_{\text{TM}} \subseteq \text{TM+State}.$$

The main goal, for this exercise, will be to prove that the language UselessState$_{\text{TM}}$ is ***unrecognizable***.

It will be useful, to get things started, to consider another language as well. Let Empty$_{\text{TM}} \subseteq \Sigma_{\text{TM}}^\star$ be the set of encodings of Turing machines $M = (Q, \Sigma, \Gamma, \delta, q_0, q_{\text{accept}}, q_{\text{reject}})$ such that $L(M) = \emptyset$. Then Empty$_{\text{TM}} \subseteq$ TM, since every string in Empty$_{\text{TM}}$ is an encoding of a Turing machine.

First Stage (Required)

It is established, in the preparatory reading for Lecture #16, that the language All$_{\text{TM}}$ is ***undecidable*** — where All$_{\text{TM}} \subseteq$ TM is the set of encodings of Turing machines

$$M = (Q, \Sigma, \Gamma, \delta, q_0, q_{\text{accept}}, q_{\text{reject}})$$

such that $L(M) = \Sigma^\star$. In particular, this is established by showing that

$$\text{A}_{\text{TM}} \preceq_{\text{M}} \text{All}_{\text{TM}}$$

using a many-one reduction $f : \Sigma_{\mathsf{TM}}^\star \to \Sigma_{\mathsf{TM}}^\star$ such that, for all $\mu \in \Sigma_{\mathsf{TM}}^\star$,

$$f(\mu) = \begin{cases} \varphi_f(\mu) & \text{if } \mu \in \text{TM+I}, \\ x_f & \text{if } \mu \notin \text{TM+I}, \end{cases}$$

where φ_f was a *mapping* from encodings of Turing machines and input strings to encodings of Turing machines, while x_f was a *fixed string* in $\Sigma_{\mathsf{TM}}^\star$.

It turns out that it is possible to show that

$$\mathsf{A}_{\mathsf{TM}}^C \preceq_{\mathsf{M}} \mathsf{Empty}_{\mathsf{TM}}$$

using a many-one reduction $g : \Sigma_{\mathsf{TM}}^\star \to \Sigma_{\mathsf{TM}}^\star$ such that, for all $\mu \in \Sigma_{\mathsf{TM}}^\star$

$$g(\mu) = \begin{cases} \varphi_g(\mu) & \text{if } \mu \in \text{TM+I}, \\ x_g & \text{if } \mu \notin \text{TM+I}, \end{cases}$$

where φ_g is also a *mapping* from encodings of Turing machines and input strings to encodings of Turing machines, while x_g is also a *fixed string* in $\Sigma_{\mathsf{TM}}^\star$.

Furthermore, one can ***either*** set φ_g to be the same mapping as φ_f ***or*** set x_g to be the same string as x_f — but not both.

1. Say whether you can choose the *mappings* (φ_f and φ_g) to be the same, for both reductions, or you can choose the *fixed strings* (x_f and x_g). Then *briefly* explain why your choice is correct, and why the other choice could not possibly be made.

2. Suppose that (as suggested here) $\mathsf{A}_{\mathsf{TM}}^C \preceq_{\mathsf{M}} \mathsf{Empty}_{\mathsf{TM}}$. What can be concluded, from this, about the language $\mathsf{Empty}_{\mathsf{TM}}$?

You may now ***assume*** that $\mathsf{A}_{\mathsf{TM}}^C \preceq_{\mathsf{M}} \mathsf{Empty}_{\mathsf{TM}}$.

3. Identify the kind of ***many-one reduction*** that should now be used to prove that the language $\mathsf{UselessState}_{\mathsf{TM}}$ is unrecognizable. That is, identify another language, L, that should be involved, and say whether you should prove that

   $$L \preceq_{\mathsf{M}} \mathsf{UselessState}_{\mathsf{TM}}$$

   or you should prove that

   $$\mathsf{UselessState}_{\mathsf{TM}} \preceq_{\mathsf{M}} L.$$

4. The many-one reduction, that you need to provide, is a ***total function*** h from the set of strings over one alphabet to the set of strings over another alphabet.

   List — in as much detail as you can — the ***properties*** that h must have, in order for it to be the kind of "many-one reduction" that you want.

   You do not need to say precisely what the function h will be, or prove that it has these properties — yet: That is what "Stage #2" of this assignment is about.

Submission Details

- Your answer for the above question will be worth 5% of your total grade for this course.
- It should not be necessary to write than three pages of text to answer these questions, and many groups may find that two (or fewer) pages are sufficient.
- Each group's answer for this question must be uploaded to the D2L dropbox for this, on or before ***11:59 pm on Wednesday, November 1***. ***No extensions will be granted for this*** and all requests to accept material, that was not successfully uploaded to D2L by this time, will be declined.

Feedback — and a Decision

Solutions for Questions #1–#3 will be available on D2L on or before 11:59 pm on Friday, November 3. Feedback for your answer for these questions will be available on D2L on or before 11:59 pm on Monday, November 6. This feedback will include comments about your answer, if it did not receive full marks, along with the answer for the following "Yes/No" question:

Can your answer for Question #3 be used to complete this assignment?

If the answer that your group receives for this question is ***No*** then your group must use the instructor's solution for Question #3 when completing the second stage of this assignment.

If the answer that your group receives for this question is ***Yes*** then your group can choose either to continue with your solution for Question #3, when completing the second stage of this assignment, or to continue with the instructor's solution instead.

Note: You may discuss your answer for Question #3 and the feedback that you got during the instructor's office hours on Tuesday, November 7. If more time for these discussions is needed then the instructor will add extra office hours for this on Wednesday, November 8.

Second Stage (Optional)

Question To Be Answered

If you did not receive full marks for your answer for Questions #1–#4 (from the "First Stage") then you may recover up to two lost marks[1] by answering the following ***optional*** question.

[1] The maximum number of marks you received for Questions #1–#4 will still be 5 out of 5.

5. As noted above, feedback that your group will receive by Monday, November 6 will be include an answer for the following "Yes/No" question:

   Can your answer for Question #3 be used to complete this assignment?

   - *If the answer that your group received for this question is **No**:* Explain *why* it would not have been possible for you to use your answer for Question #3 to prove that the language, $\text{UselessState}_{\text{TM}}$, is unrecognizable. What should you do in order to avoid making the same kind of mistake again?
   - *If the answer for Question #3 was **"Yes"** but you did not receive full marks for Questions #1–#3 for some other reason:* Explain why you lost marks and describe something specific that you could do, ***possibly involving the use of resources that are available to you at the University of Calgary***, to improve your work and avoid losing marks for the same reason in the future

   Your answer should include more than the marker's feedback for your answer: Try to show that you understand whatever mistake you made and what you might do, differently, if you were asked to prove the undecidability or unrecognizability of another language, later on in this course.

   It might be helpful to be thinking about this question — and developing an answer for it — while you are developing answers for Questions #6–#8, below. Where would you get stuck when you try to answer these questions, when starting with your answer for Question #3?

Submission Details

- It should not be necessary to write than one page of text to answer this question, and many groups may find that one half-page is sufficient.
- Each group's answer for this question must be uploaded to the D2L dropbox for this, on or before ***11:59pm on Friday, November 10. No extensions will be granted for this*** and all requests to accept material, that was not successfully uploaded to D2L by this time will be declined.

Second Stage (Required)

6. Say whether your group is completing the rest of this assignment by starting with its answer for Question #3 (instead of using the instructor's solution). *This will not be worth any marks but will be used by a marker when the rest of your work is considered.*

At this point you have identified languages $L_1 \subseteq \Sigma_1^\star$ and $L_2 \subseteq L_2^\star$, such that you are going to prove that

$$L_1 \preceq_{\mathsf{M}} L_2.$$

7. Prove that $L_1 \preceq_{\mathsf{M}} L_2$.

   Note: Lectures #16 and #17 present and use a *process* that can be used to discover proofs of claims, like the above one, and to *organize* the proof. ***It is strongly recommended that you follow this process*** — it should make this problem more manageable than it otherwise would be.

8. The lecture presentation for Lecture #17 introduced another language $\mathsf{E}_{\mathsf{TM}} \subseteq \Sigma_{\mathsf{2TM}}^\star$ and provided a proof that E_{TM} is ***undecidable***.

   Say — reasonably briefly — how you could prove that the language E_{TM} is ***unrecognizable***, as well.

   In particular, identify languages $L_3 \subseteq \Sigma_3^\star$ and $L_4 \subseteq \Sigma^\star$ (and alphabets Σ_3 and Σ_4) such that you would prove this result by showing that

   $$L_3 \preceq_{\mathsf{M}} L_4$$

   and describe, briefly, how to ***modify*** one of the many-reductions that you will have seen, in this course, by Lecture #17, in order to establish this.

Submission Details

- Answers for the Questions #6–#8 with be worth 10% of your total grade for this course.
- The number of pages needed to answer Questions #6–#8 will depend on several things — including how much time you spend on *organizing* your answers (and how good your organization was). Well-organized answers can often just be just as "complete" as poorly organized ones and can be significantly shorter, at the same time.

  Furthermore, you can sometimes discover a better way to write something (that results in a clearer *and* shorter answer) while you are writing something up. This is one of the reasons why you should always try to allow as much time for writing as you can: Multiple drafts of your answer might be needed before you produce one that you can be proud of.
- ***A general extension has been requested, and granted:*** Your group's answers for these questions must be uploaded to the D2L dropbox for this, on or before ***11:59 pm on Tuesday, November 14,*** in order to be considered to be on time.

  Late submissions for the "required" part of the second stage will be accepted with a penalty. In particular, 10% of the total marks for available will be subtracted from the mark you would receive, if your submission is late but is submitted by 11:59 pm on Wednesday, November 15.[2] If your submission is later than that, then 25% of the total marks will be available will be subtracted from the mark that you would receive, if your submission is submitted by 11:59 pm on Thursday, November 16. No submissions will be accepted after that. All work must be submitted using the appropriate dropbox on D2L.

All requests to accept material that was not successfully uploaded to D2L will be declined.

[2] However, the mark you receive will never be less than zero.