Programming with Induction, Recursion, and Algebraic Laws

COMP 105 Assignment

Due Tuesday, January 29, 2019 at 11:59PM

Contents

This first assignment is divided into three parts:

- *Comprehension questions* that help you focus your reading
- *Programming exercises* that reinforce your programming process and that develop skills with induction and recursion
- A *photograph* for the private use of the course staff, which will help us learn to recognize you and call you by name (and which you will have submitted early)

The assignment is all individual work. There is no pair programming.

NOTE: This assignment is due one minute before midnight. You may turn it in *up to 24 hours after the due date*, which will cost you one extension token[1]. If you wish not to spend an extension token, then when midnight arrives submit whatever you have. We give partial credit.

ALERT: This assignment is significantly easier than a typical COMP 105 assignment. Its role is to get you acclimated and to help you start thinking systematically about writing code from scratch, especially using algebraic laws to write recursive functions. Later assignments get harder and more time-consuming, so don't use this one to gauge the difficulty of the course.

Reading-Comprehension Questions (10%)

Before starting the programming problems, answer the following questions about the reading. Place your answers in a plain text file called `cqs.impcore.txt`. The best procedure is to download the questions[2] and edit in your answers.

Please read pages 6–14 in *Programming Languages: Build, Prove, and Compare*.

1. What is the value of the following Impcore expression?

   ```
   (if (> 3 5) 17 99)
   ```

2. Which of the following best describes the syntactic structure of Impcore?

   (a) An expression can be made with any definition as one of its parts
   (b) A definition can be made with any expression as one of its parts
   (c) Both of the above
   (d) None of the above

3. Does the following Impcore test pass? Please answer "yes" or "no."

   ```
   (check-expect (+ 1 2 3) 6)
   ```

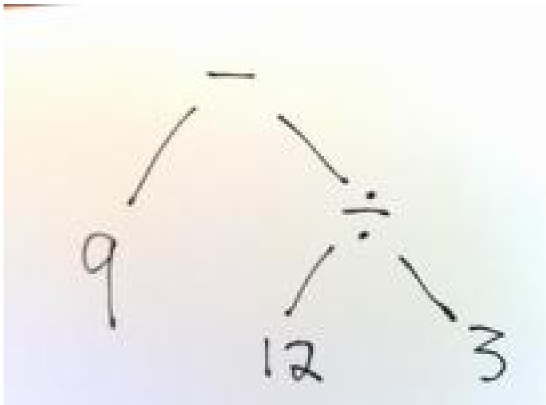
   Assuming `x` is bound to a global variable, does the following Impcore test pass? Again, please answer "yes" or "no."

   ```
   (check-expect (set x 1) 1)
   ```

Next read section 1.2, which starts on page 15, about abstract syntax.

4. After reading about abstract syntax, look at this picture of an abstract-syntax tree for a "calculator expression":

[1] ../syllabus.html#what-if-i-cant-get-my-homework-in-on-time
[2] ./cqs.impcore.txt



Answer these questions:

(a) What concrete syntax could you write in C for this expression?
(b) What concrete syntax could you write in Impcore for this expression?

In *Seven Lessons in Program Design*, at https://www.cs.tufts.edu/comp/105/design/lessons.pdf, read the Introduction and Lesson 1.

5. I show you a recursive function `f` that takes one argument, a natural number `n`. The structure of `n`, and therefore the internal structure of `f`, are based on the Peano proof system from the handout.

   (a) What are the different ways `n` can be formed?
   (b) When `f` is given `n`, what code do you expect `f` to use to determine how `n` was formed?
   (c) For which values of `n` do you expect `f` to make a recursive call?
   (d) When a recursive call is made, what value is passed as argument?

Read the expectations about contracts in the *course coding guidelines*[3].

6. Suppose I write a contract for a `power` function that says, "this function multiplies x by itself n times." According to our expectations, is this a good contract or a bad contract? Please answer "good" or "bad."

7. In *Seven Lessons in Program Design*, just before the end of Lesson 1, you will find a section on "Complete process examples." This section suggests that the `factorial` function—but not the `power` function—could be submitted without a contract.

   (a) Why would it be OK to submit the `factorial` function without a contract? For an idea, look at the "Exemplary" column in the "Documentation" section of the general coding rubric.
   (b) Why doesn't the same argument apply to the `power` function? For an idea, check the design lesson.

Programming in Impcore (80%)

The problems below serve multiple purposes:

- To give you practice applying the software process presented in the first lesson on program design, including the forms of testing and documentation that are expected in the course

[3] ../coding-style.html#contracts

- To get you thinking explicitly about induction and recursion
- To get you acclimated to the LISP-style concrete syntax used throughout the course
- To get you started with the course software

You can start these exercises immediately after the first lecture. You may write very efficient solutions—but do not feel compelled to do so. Just make sure that your recursive functions terminate!

Do not share your solutions with anyone. We encourage you to discuss ideas, but **no other student may see your code**.

Getting set up with the software

The interpreter you need to run Impcore code is provided as part of the course. To add course software to your execution path, run

```
use -q comp105
```

You may want to put this command in your `.cshrc` or your `.profile`. The `-q` option is needed to prevent `use` from spraying text onto standard output, which may interfere with `scp`, `ssh`, `git`, and `rsync`.

The `impcore` interpreter is in `/comp/105/bin`; if you have run `use` as suggested above you should be able to run it just by typing

```
ledit impcore
```

The `ledit` command gives you the ability to retrieve and edit previous lines of input.

If your code and unit tests are in file `solution.imp`, you can load and run them by typing

```
impcore -q < solution.imp
```

Unit testing

The special "extended-definition forms" `check-expect`, `check-assert`, and `check-error` are part of every language in the book. For example, as described in section 1.1.1 of the book (page 6), they are part of the Impcore language. These forms serve both as unit tests and as documentation. **Every function you write must be tested and documented using `check-expect` or `check-assert`**, and possibly also `check-error`. The number of unit tests must be appropriate to the function's contract and to the structure of its input, as described in the first lesson on program design.

How do you know when to use `check-expect` and when to use `check-assert`? Like this: `check-expect` tells if two expressions evaluate to the same value, and `check-assert` checks to see if a single expression evaluates to truth or falsehood. To be confident you are getting things right,

- If you are writing a `check-expect` that something is 0 or 1, meant as a Boolean, then probably you should be writing `check-assert`. (If checking for 0, use `check-assert` with `not`, as shown below.)

- If you are writing a check-assert of an expression written with `=`, you should *definitely* write `check-expect` instead. It's more readable, and if it fails, you get a more useful error message.

Expectations for your software-design process

To write code from scratch, we recommend the nine-step design process that is presented in the introduction to *Seven Lessons on Program Design*[4]. In this homework, the process is primarily for practice—many students will be able to do the problems without the help that the process provides. But in a couple of weeks, you will need to have a good process, or you will find that the homeworks take too much time.

Our process has nine numbered steps. On this homework, here's what we expect for each step:

1. All data are integers, and most are natural numbers. We expect you to be able to break natural numbers down into the forms described in the first lesson on program design. Those forms will appear on the left-hand sides of the algebraic laws you write for step 6.

2. We expect you to be able to choose example inputs, which we expect to see in your unit tests.

3. We expect you to choose names only for any helper functions that you choose to write. Most problems, but not all, can be solved without helper functions. When there are helper functions, they will be scrutinized carefully, and their names will be judged according to the general coding rubric[5].

4. We expect each function you submit to be preceded by a short contract that appears in a comment.

    - Contracts for the assigned functions don't require much thought; language from the book (or this homework) is fine.

    - Contracts for helper functions are another story. For each helper function you define, *choose a good name*, and write a *specific, accurate contract.* Use the coding guidelines.[6] Your contract must meet the expectations laid out in the Documentation section of the general coding rubric[7].

    The contracts for the helper functions will weigh more heavily toward your grade.

    Here's a model contract:

    ```
    ;; (has-digit? n d) returns 1 if and only if the decimal
    ;; representation of positive integer `n` includes a decimal
    ;; digit that is equal to `d`; it returns 0 otherwise.
    ```

5. We expect unit tests to be submitted with each function, and we expect them to *follow* the function and to be indented eight spaces. We expect only the basics: one or two unit tests for each form of input.[8] Additional unit tests are acceptable, but they must be separated from the basic tests by a blank line.

    Here are two example unit tests:

    ```
    (check-assert (not (has-digit? 123 7)))
    (check-assert      (has-digit? 123 2))
    ```

6. We expect you to write algebraic laws for any function that contains an `if` expression or a recursive call. Algebraic laws should appear between a function's contract and its definition.

[4] ../design/lessons.pdf

[5] ../coding-rubric.html

[6] ../coding-style.html#contracts

[7] ../coding-rubric.html

[8] For ordinary results, we expect one test per form of input. For Boolean results, we expect one "true" test and one "false" test for each form of input, when possible.

7. We expect the body of each function to have at most one case per algebraic law.
8. We expect the body of each function to be consistent with the function's algebraic laws.
9. We expect each function's unit tests to be indented eight spaces, and to pass. When a function is meant to return only 1 or 0, coding for "true" or "false," we expect that function's unit tests to use `check-assert` and to test *both* outcomes.

Help with the process: a solution template

To help you organize and present the results of your work, we provide a solution template[9]. Copy the template into file `solution.imp`. The template provides placeholders for contracts (design step 4), unit tests (step 5), algebraic laws (step 6), and function definitions (steps 7 and 8).

In the template, you will see that most functions are followed by a single unit test that uses `check-error`. The test is a placeholder. Remove the `check-error` and replace it with unit tests that use `check-expect` or `check-assert`, which you must write. (For most functions, you will need multiple unit tests: at least one per algebraic law.)

Indent all unit tests by 8 spaces.

A placeholder function definition has body `(/ 1 0)`. Evaluating this code divides 1 by 0, which causes an error. The solution template should pass all unit tests, as reported by

```
impcore -q < solution.imp
```

The template does not include placeholders for helper functions. If you write any helper functions, *supply each helper function with a contract, laws, and unit tests.*

Quick help with Impcore

The concrete syntax of Impcore can be found on page 6 of *Programming Languages: Build, Prove, and Compare*. As a quick summary, the following code uses every possible syntactic form of expression. But on this assignment, **you must not use the `while` and `set` forms**, and the code you *submit* must not print.

```
(define even? (n) (= (mod n 2) 0))

(define 3n+1-sequence (n) ; from Collatz
  (begin
    (while (!= n 1)
       (begin
         (println n)
         (if (even? n)
             (set n (/ n 2))
             (set n (+ (* 3 n) 1)))))
    n))
```

The "initial basis" of a programming language is a fancy technical name for the names that are already defined. In Impcore, the initial basis comprises primitive and predefined functions, which have these names:

[9] http://www.cs.tufts.edu/comp/105/homework/solution-template.imp

```
mod      or       println  /
!=       and      =        *
>=       printu   >        -
<=       print    <        +
not
```

The problems you must solve

You will solve three groups of ordinary problems and one challenge problem. Do the problems in the order in which they appear below, not in book-numbering order.

Direct applications of the design method

Each problem in the first group can be solved by direct application of the methods sketched in the first lesson on program design: choose a proof system, write algebraic laws, design the code.

The problems are as follows:

- **L10.** Define a function `log10` that approximates the base-10 logarithm. More exactly, when $n > 0$, `(log10` n`)` returns the smallest integer k such that $10^{k+1} > n$. This problem is a special case of the `log` function from exercise 5.
- **DD.** Define a function `double-digit`. When given a positive integer less than 20,000, `double-digit` returns a positive integer whose decimal representation is the same as the decimal representation of the input, except each digit appears twice. For example, `(double-digit 123)` is `112233`.
- **C.** Define a function `population-count`, which when given a nonnegative integer, returns the number of 1 bits in the binary representation of that integer. (This function is named for a machine instruction found on Intel CPUs.)
- **10**. Define the function `binary` described in exercise 10 on page 78 of *Build, Prove, and Compare*.

Generalization of the design method

Each problem in the second group can be solved by generalizing the methods in the first lesson on program design.

- **5.** Define the function `log` that is specified in exercise 5 on page 76 of *Build, Prove, and Compare*. You will have to come up with a new proof system for natural numbers, which should generalize the ideas in the lesson. You will be able to write algebraic laws using this new proof system. Following your laws, place a comment sketching the new proof system. Sketch the proof system using informal English, as in this example from the Seven Lessons[10]: "A natural number is either zero or is ten times a natural number m plus a decimal digit d."
- **4.** Define the function `sigma` that is specified in exercise 4 on page 76 of *Build, Prove, and Compare*. Although m and n are not guaranteed to be natural numbers, this problem specifies $m \leq n$, so the difference $n - m$ is a natural number. Your algebraic laws should handle cases for $n - m = 0$ and $n - m = k + 1$, where k is a natural number. The laws should be written like this:

[10] ../design/lessons.pdf

```
;; laws:
;;   (sigma m m) == ...
;;   (sigma m n) == ..., where n > m
```

If you make a recursive call for the case $n-m = k+1$, the recursive call should satisfy $n-m = k$.

Recursion that is not structural

A recursive computation that is driven by the structure of the data, as in the first two groups of problems, is called *structural recursion*. Not all recursions are structural.

- **8**, part one. Define the function `prime?` that is specified in exercise 8 on page 76 of *Build, Prove, and Compare*. I recommend searching for a divisor. This search will require a recursive helper function, which is not structural.

  All helper functions must begin with algebraic laws. When the recursion is not structural, algebraic laws do not discriminate on the *forms* of the input. Instead, they discriminate on *properties* of the input. For this problem, one interesting property is that one number evenly divides another. There is also one other interesting property. In each law, write the interesting properties for that law as a *side condition*, which should appear after the right-hand side of the law, following a comma and the word "where." As a model, look at the side condition for the property $n > m$ in the laws for `sigma` above.

  For the function `prime?` itself, I recommend discriminating using three mutually exclusive properties: either $n < 2$, $n = 2$, or $n > 2$. (Function `prime?` expects a nonnegative integer.)

A challenge problem

- **8**, part two. Define the function `nthprime` that is specified in exercise 8 on page 76 of *Build, Prove, and Compare*. This one should be implemented using structural recursion, and I recommend a structurally recursive helper function. The challenge here is to come up with a good name and a clear contract for the helper function.

  Note that function `nthprime` expects a *positive* integer.

Other expectations for your solutions

For this assignment, we expect you to apply the recommended design process and to deliver working functions with good names, clear contracts, algebraic laws, and unit tests. We also expect the following:

- Your solutions **must be valid Impcore**; in particular, they must pass the following test:

  `/comp/105/bin/impcore -q < solution.imp > /dev/null`

  with no error messages and no unit-test failures. If your file produces error messages, we won't test your solution and you will earn No Credit for functional correctness (you can still earn credit for readability).

- Your solutions must load and complete within 250 CPU milliseconds. If you write long-running tests, don't include them in `solution.imp`; instead, create a file `extra-tests.imp`. This file should include a line `(use solution.imp)`.

- On this assignment, as on several assignments to come, **your code must use recursion.** Code using `while` loops will receive No Credit.

- Code you submit must not call `print`, `println`, or `printu`. Code that prints is likely to fail our automated tests.
- You may use helper functions where appropriate, but **your code must not define or use global variables**.
- Your code must be your own work. **Do not share your solutions with anyone.** We encourage you to discuss ideas, but **no other student may see your code**.
- Your algebraic laws must be your own work. Algebraic laws have the same status as code.

A photograph we can use to learn your name (10%)

The submit scripts for this assignment will look for the photograph you will have already submitted.

What to submit and how to submit it

You'll put your assignment in the same directory in which you have already submitted your photograph. In that directory, you will create or copy these files:

- `cqs.impcore.txt` will contain the comprehension questions and your answers.
- `solution.imp` will contain your code, with its documentation and unit tests. Problems will appear in the order in which they are listed above, which is also the order of the solution template[11].
- `photo.jpg` will contain a recognizable photograph of you. This can be a copy of the photo you *already* submitted.
- `README` will identify anyone with whom you have collaborated or discussed the assignment and will include any other information you wish to pass on to us. We provide a `README` template[12]; please use it.

As soon as you have the files listed above, run `submit105-impcore` to submit a preliminary version of your work. The submit script checks your work and runs `provide` on your behalf. Do submit early, then keep submitting until your work is complete; we grade only the last submission.

The submit script will ask you some questions, the answers to which may be carried over from your photograph submission:

- Your preferred first and last names
- How we should pronounce your name, as in "kaeth-LEEN FI-shur" or "NORE-muhn RAM-zee."
- How many hours you worked on the assignment

You may also submit `extra-tests.imp`, which should contain only test code, unit tests (`check-expect` or `check-error`), and the line `(use solution.imp)`. You can run the tests using the Unix command

~~cat solution.imp extra-tests.imp | impcore -q~~

```
impcore -q < extra-tests.imp
```

[11] solution-template.imp
[12] README_template

How your work will be evaluated

How your code will be evaluated

In this assignment, you learn how we expect your code to be structured and organized. Our expectations are presented on the coding-style page[13] on the course web site. When we get your work, we will evaluate it in two ways:

- About 50% of your grade for the assignment will be based on our judgement of the structure and organization of your code. To judge structure and organization, we will use the following dimensions:
  - *Documentation* assesses whether your code is documented appropriately.

    We expect you to document each function such that someone else could use your code and reason about its result without having to see the code itself. In particular, every function must be documented with a contract, and the contract must mention each parameter by name.
  - *Algebraic laws* are also documentation, but they serve primarily as a design tool. Your algebraic laws must demonstrate your understanding of the forms of input: we expect one law per form of input. If any form of input requires more than one law, we expect the laws to be disambiguated with mutually exclusive side conditions.

    We expect the laws to be the foundation for your code: your code must begin by examining inputs to determine which algebraic law governs the given input.

    Write your algebraic laws before you start coding. In the first half of the course, this "design first" practice might save you forty or fifty hours.
  - *Unit testing* assesses whether your code is appropriately tested by some combination of `check-expect`, `check-assert`, and/or `check-error`. Appropriate testing exercises every form of input and every part of your code. We expect a test for every algebraic law.
  - *Form* assesses whether your code uses indentation, line breaks, and comments in a way that makes it easy for us to read.

    We expect you to use consistent indentation, to obey the offside rule[14] described in the coding-style guidelines, and to limit the use of inline comments in the body of each function.

    We expect you to indent each block of unit tests by eight spaces.
  - *Naming* assesses your choice of names. To people who aspire to be great programmers, names matter deeply.

    We give you a hand here by providing a template in which the names of top-level functions and their arguments are already chosen for you. For helper functions, you will choose your own names. Look at the general coding rubric[15] and choose wisely.
  - *Structure* assesses the underlying structure of your solution, not just how its elements are documented, formatted, and named.

[13] ../coding-style.html
[14] ../coding-style.html#offside
[15] ../coding-rubric.html

We expect that your solutions will use recursion and function calls, not loops and assignments. Additionally, we expect the case structure of each function to follow that function's algebraic laws.

- About 30% of your grade for the assignment will be based on our judgement of the correctness of your code. We often look at code to see if it is correct, but our primary tool for assessing correctness is by testing. On a typical assignment, the correctness of your code would carry more weight, but this assignment is all about following the recommended software process and using algebraic laws effectively, so test results carry less weight.

The detailed criteria[16] we will use to assess your code are found at http://www.cs.tufts.edu/comp/105/coding-rubric.html. Though things may be worded differently, most of these criteria are also applied in COMP 11, 15, and 40—they make explicit what we mean by "good programming practice." But as you might imagine, there is a lot of information there—probably more than you can assimilate in one reading. The highlights are

- *Documentation*
  - Each function is documented with a contract that explains what the function returns, in terms of the parameters, which are mentioned by name. From documentation, it is easy to determine how each parameter affects the result.
  - The contract makes it possible to use the function without looking at the code in the body.
  - If not all inputs are permissible, the contract determines what inputs are and are not OK.
  - The contract appears consistent with the laws, code, and unit tests.
  - Each parameter is mentioned in its function's contract.
- *Form*
  - Code fits in 80 columns.
  - Code respects the offside rule.
  - Code contains no tab characters.
  - Indentation is consistent everywhere.
  - If a construct spans multiple lines, its closing parenthesis appears at the end of a line, possibly grouped with one or more other closing parentheses—never on a line by itself.
  - No code is commented out.
  - Solutions load and run without calling `print`
- *Naming*
  - Each function is named either with a noun describing the result it returns, or with a verb describing the action it does to its argument. (Or the function is a predicate and is named as suggested below.)
  - A function that is used as a predicate (for if or while) has a name that is formed by writing a property followed by a question mark. Examples might include even? or prime?. (Applies only if the language permits question marks in names.)
  - In a function definition, the name of each parameter is a noun saying what, in the world of ideas, the parameter represents.
  - Or the name of a parameter is the name of an entity in the problem statement, or a name from the underlying mathematics.
  - Or the name of a parameter is short and conventional. For example, a magnitude or count might be n or m. An index might be i, j, or k. A pointer might be p; a string might be s. A variable might be x; an expression might be e.
- *Laws*

[16] ../coding-rubric.html

  - Each left-hand sides applies the defined function to forms of input.
  - Laws are distinguished by distinct forms of input, or (only when necessary) by mutually exclusive side conditions.
  - Every permissible input is handled by exactly one law, and no forms of input are omitted.
  - Each law specifies a result that is consistent with the function's contract, and it does not say anything about inputs that are forbidden by the contract.
  - The right-hand side of a law refers only to those variables that appear on the left-hand side of that law.
  - Variables are used consistently between the left-hand side and the right-hand side of each law.
- *Structure*
  - Solutions are recursive, as requested in the assignment.
  - Any case analysis begins by identifying which law applies to the input.
  - There's only as much code as is needed to do the job.
  - In the body of a recursive function, the code that handles the base case(s) appears before any recursive calls.
  - The code of each function is so clear that, with the help of the function's contract, course staff can easily tell whether the code is correct or incorrect.
  - Expressions cannot easily be simplified.
- *Unit Testing*
  - Every test uses inputs whose behavior is constrained by contract.
  - There is a unit test for each form of input.
  - There is one unit test for each algebraic law (should overlap with input tests).
  - If a function returns a Boolean, its tests are written using `check-assert` and possibly `not`.
  - If a function returns a Boolean, then for each form of input that can return either true or false, there is both a truth test and a falsehood test.

How your photograph will be evaluated

The evaluation of your photograph is described in the photo assignment[17].

[17] photo.html#how-your-photograph-will-be-evaluated