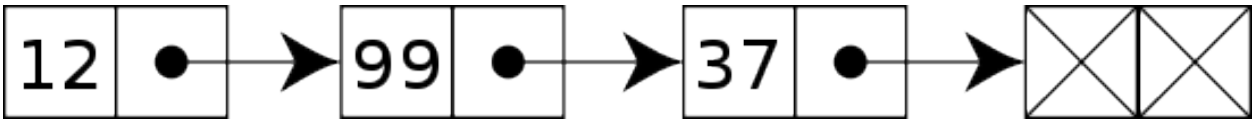
15-122: Principles of Imperative Computation

Recitation 7 — Josh Zimmerman

Linked lists

A linked list is a data structure that allows you to easily store a variable amount of data in a list.



Note that by convention we will normally use a *sentinel* (or *dummy*) node in 122 that flags the end of the list. It is uninitialized. When we see that sentinel node, we know we're at the end of the linked list.

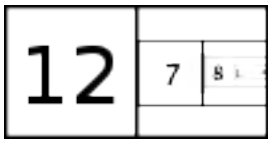
Adding a node to the linked list is as simple as initializing the data in the sentinel node, making a new sentinel node, and pointing the old sentinel node to the new sentinel node.

In C0, we can define a node in a linked list using a `struct` datatype (in this case, the linked list we make stores `int`s):

```
struct list_node {
    int data;
    struct list_node* next;
};
```

A `struct` is simply a way of defining a datatype that is an aggregation of other datatypes. In this case, we're saying that a `struct list_node` is a datatype that has an `int` and a pointer to another `struct list_node`. Note that it's crucial that we have a pointer to another `struct list_node` rather than a `struct list_node`: if we had a `struct list_node` instead, we'd have a `struct list_node`-ception (to be more formal, we'd recurse infinitely in our definition of the `struct list_node`): in every `struct list_node`, we'd have another `struct list_node`.

For a more concrete explanation of why that alternate definition is bad, see this image:



The reason that this doesn't happen with a `struct list_node*` is that if we have a pointer to a `struct list_node` all that this does is say "this variable (`next`) will either be `NULL` or tell us *where to find* a `struct list_node`."

Pointers, or "Analogies are like your car. They can break down if you're not careful about taking it to the mechanic"

At this point, we need to go into more depth about what exactly a pointer is.

In C0, for every datatype, there is a corresponding *pointer* that can point to that datatype. For example, for `int`s, there is an `int*` datatype. We know that anything of type `int*` will either be `NULL` or tell us where to find an `int`.

You can imagine a pointer as sort of like a physical address. If I have the address of my favorite restaurant saved on my computer, then that tells me where I can go to find it. If the restaurant goes out of business and gets replaced with a different restaurant, I'll still have the address saved on my computer and I can still go to that address, I just won't find what I expect to. Similarly, if a demolition crew also has a copy of the address of my favorite restaurant and it goes there and knocks the building down, then the building will be knocked down when I next visit it later that day.

A pointer is just the address of a location in your computer's memory and works similarly. If someone else has a copy of that address and they modify the part of memory that the pointer points to (or, in other words, that is at that address), then you'll see that modification the next time you visit that address.

Note that if you want to look at what is at the address that a pointer talks about, you need to do use the `*` operator. If `p` is a pointer, `*p` says "go to the address that `p` specifies and bring me what is there." This is referred to as *dereferencing* a pointer.

Since C0 requires that we pass around pointers to `structs` rather than the `structs` themselves (and since it's good normally good practice to do that in C as well), there's some syntax that we use when we want to both dereference a pointer to a `struct` and access a field of that struct: `(*p).a` will dereference `p` and get the field in it that's called `a`, and so will `p->a`. These two functions do the same thing (but the one on the right is easier to read).

```
void incr_data(struct list_node* l)
//@requires l != NULL;
{
    (*l).data = (*l).data + 1;
}
```

```
void incr_data(struct list_node* l)
//@requires l != NULL;
{
    l->data = l->data + 1;
}
```

A `NULL` pointer is sort of like an address of a house that someone else lives in and that is guarded by violent hungry alligators. If you attempt to go to it, the alligators will eat you (unless you are already authorized to go to the house, which you won't be, at least until you take OS, and maybe not even then).

In C0, following a `NULL` pointer will crash your program with a segmentation fault. For this reason, it's critical to *always* make sure that any time you follow (or *dereference*) a pointer that that access is safe, just like what you must do when accessing an array. You should add contracts and conditional statements to your code to ensure that any pointer access is safe. For instance, the code fragment on the left is BAD because it could cause a segmentation fault, but the code segment on the right will work assuming that its preconditions are met. (Note that neither of these code fragments modify the data that `p` points to since there are no assignments.)

BAD – DO NOT EVER DO

```
int follow_and_add_one(int* p)
{
    // If p is NULL this crashes
    return *p + 1;
}
```

Good

```
int follow_and_add_one(int* p)
//@requires p != NULL;
{
    return *p + 1;
}
```

There will be some cases that you *cannot* use contracts like the above to guarantee you aren't holding a

NULL pointer. In those cases, you should instead use conditional statements to guarantee that you won't dereference NULL.

It's **really, really important** to note that the analogy of physical addresses and doesn't work as well when you examine it in depth. I'm using it because it can help you to understand the general concept behind a pointer. Later in the semester, we'll examine pointers in more depth and see how they work in C, but for now the main important takeaways are:

- A pointer tells you *where* to find something. If that something changes, you can still look where the pointer tells you to look, but you'll find something different than what you originally put there.
- If p is a pointer that is not NULL, *p *dereferences* p, meaning it lets us directly read and modify the data that is stored at the place where p points.
- If sp is a non-NULL pointer that points to a struct, then sp->data means the same thing as (*sp).data.
- Dereferencing NULL pointers is an error and will cause your program to crash in C0. Whenever you dereference a pointer, you should have proof that it is not NULL.

alloc

Now we know what a pointer is, but how do we actually *get* one?

The expression alloc(t) will give us a pointer to an area of memory that can hold something of type t. For example, if I write int* a = alloc(int); then the computer will make space somewhere in memory for an int and will put the address of that area of memory in a. In other words, after we execute that expression, a will be a pointer that points to an int.

You can use this for any type: If I want space for a node in a linked list, I'd type struct list_node* list = alloc(struct list_node); and then list would be a pointer to an struct list_node.

Practice!

(Credit for this section goes to CMU alumna Caroline Buckey; it has been updated since by Alex Cappiello and Rob Simmons.)

Suppose you have the implementation using linked lists shown in lecture. Specifically, you have the following structs:

```
1  struct list_node {
2    int data;
3    struct list_node* next;
4  };
5  typedef struct list_node list;
6
7  struct linkedlist_header {
8     list* start;
9     list* end;
10 };
11 typedef struct linkedlist_header linkedlist;
```

In lecture, we talked about the is_segment(start, end) function that tells us we can start at start, follow next pointers, and get to end without ever encountering a NULL. (We won't worry about the

problems with getting `is_segment` to terminate in this recitation.) A `linkedlist` is a non-`NULL` pointer that captures a reference to both the start and end of a linked list.

```
1 bool is_linkedlist(linkedlist* L) {
2   if (L == NULL) return false;
3   return is_segment(L->start, L->end);
4 }
```

Recall from lecture that we always have one "dummy" node at the end of our linked list segments. Its fields are uninitialized; it simply ensures that we never need to worry about `start` or `end` being null.

Creating a new linked list

Here's the code that creates a new linked list with one non-dummy node. Suppose `linkedlist_new(12)` is called. For each of lines 4-9 (inclusive) draw a diagram that shows the state of the linked list after that line executes. Use `X` for struct fields that we haven't initialized yet.

```
1 linkedlist* linkedlist_new(int data)
2 //@ensures is_linkedlist(\result);
3 {
4   list* p = alloc(struct list_node);
5   p->data = data;
6   p->next = alloc(struct list_node);
7   linkedlist* L = alloc(struct linkedlist_header);
8   L->start = p;
9   L->end = p->next;
10  return L;
11 }
```

4.

5.

6.

7.

8.

9.

Adding to the end of a linked list

We can add to either the start or the end of a linked list. The following code adds a new list node to the *end*.

```
1 void add_end(linkedlist* L, int x)
2 //@requires is_linkedlist(L);
3 //@ensures is_linkedlist(L);
4 {
5   list* p = alloc(struct list_node);
6   L->end->data = x;
7   L->end->next = p;
8   L->end = p;
9 }
```

Suppose `add_end(L, 3)` is called on a linked list `L` that contains before the call, from start to end, the sequence `(1, 2)`. Draw the state of the linked list after each of lines 5 - 8 (inclusive). Include the list struct separately before it has been added to the linked list.

5.

6.

7.

8.

Adding to the start of a linked list

With the previous example in mind, can you think about what code would be necessary if we instead wanted to add a new list node to the *start* of a linked list?

```
void add_start(linkedlist* L, int x)
//@requires is_linkedlist(L);
//@ensures is_linkedlist(L);
{









}
```

Removing the first item from a linked list

This is the code that removes the first element from a linked list. If it were not for the second precondition, we might remove the dummy node! This would almost certainly cause the postcondition to fail.

```
1 int remove(linkedlist* L)
2 //@requires is_linkedlist(L);
3 //@requires L->start != L->end;
4 //@ensures is_linkedlist(L);
5 {
6   int x = L->start->data;
7   L->start = L->start->next;
8   return x;
9 }
```

Suppose `remove(L)` is called on a linked list `L` that contains before the call, from start to end, the sequence `(4, 5, 6)`. Draw the state of the linked list after lines 6 and 7 execute. Include an indication of what data the variable `x` holds.

6.

7.