Admin

Assignment 7

Learning communities: may attend any group going forward

Grading update
- 5 graded
- 4.1 soon!
- Checkpoint revisions graded

Checkpoint 2

2 pages of notes

From hashtables (9/2) through dynamic programming (10/13)

Practice problems available in group assignment 7

A problem

Input: a number k

Output: \( \{n_p, n_n, n_d, n_q\} \), where

\[ n_p + 5n_n + 10n_d + 25n_q = k \]

and \( n_p + n_n + n_d + n_q \) is minimized

What is this problem?
How would you state it in English?
Making change!

Input: a number k

Output: \( \{n_p, n_n, n_d, n_q\} \), where \( n_p + 5n_n + 10n_d + 25n_q = k \) and \( n_p + n_n + n_d + n_q \) is minimized

Provide (U.S.) coins that sum up to \( k \) such that we minimize the number of coins

Algorithm to solve it?

Making change!

Input: a number k

Output: \( \{n_p, n_n, n_d, n_q\} \), where \( n_p + 5n_n + 10n_d + 25n_q = k \) and \( n_p + n_n + n_d + n_q \) is minimized

Algorithm vs heuristics

What is the difference between an algorithm and a heuristic?

Algorithm: a set of steps for arriving at the correct solution

Heuristic: a set of steps that will arrive at some solution
Making change!

\[ n_q = \lfloor k/25 \rfloor \quad \text{pick as many quarters as we can} \]

Solve:
\[ n_q + 5n_d + 10nd = k - 25\lfloor k/25 \rfloor \quad \text{recurse} \]

Algorithm or heuristic?

Need to prove its correct!

Greedy algorithms

What is a greedy algorithm?

Algorithm that makes a local decision with the goal of creating a globally optimal solution

Method for solving problems where optimal solutions can be defined in terms of optimal solutions to sub-problems

What does this mean? Where have we seen this before?

Greedy vs. divide and conquer

Divide and conquer

To solve the general problem:

Break into sum number of sub problems, solve:

then possibly do a little work

Divide and conquer

To solve the general problem:

The solution to the general problem is solved with respect to solutions to sub-problems!
Dynamic programming

To solve the general problem:

The solution to the general problem is solved with respect to solutions to sub-problems that overlap!

Greedy vs. divide and conquer

Greedy

To solve the general problem:

Pick a locally optimal solution and repeat

The solution to the general problem is solved with respect to solutions to sub-problems!

Slightly different than divide and conquer

Proving greedy algorithms correct

One approach, prove:

1) Optimal substructure: The optimal solution contains within it the optimal solution to subproblems

2) Greedy choice property: The greedy choice is contained within some optimal solution
Making change!

\[ n_q = \lfloor k / 25 \rfloor \] pick as many quarters as we can

Solve:
\[ n_p + 5n_d + 10n_d = k - 25\lfloor k / 25 \rfloor \] recurse

\{c_1, c_2, c_3, \ldots, c_m\} solution: individual coins selected

Optimal substructure

If \(\{c_1, c_2, c_3, \ldots, c_m\}\) is optimal for \(k\), then
\(\{c_2, c_3, \ldots, c_m\}\) is optimal for \(k-c_1\)

We can combine a greedy choice with the optimal solution for the remaining problem and get a solution to the general problem

Optimal substructure

Proof by contradiction:

Assume \(\{c_1, c_2, c_3, \ldots, c_m\}\) is optimal for \(k\), but \(\{c_2, c_3, \ldots, c_m\}\) is not optimal for \(k-c_1\)

What does that imply?

Optimal substructure

Proof by contradiction:

Assume \(\{c_1, c_2, c_3, \ldots, c_m\}\) is optimal for \(k\), but \(\{c_2, c_3, \ldots, c_m\}\) is not optimal for \(k-c_1\)

There is some other set of coins \(\{c'_2, c'_3, \ldots, c'_n\}\) where \(n < m\) that add up to \(k-c_1\)

Any problem contradiction?
Assume $c_1, c_2, c_3, ..., cm$ is optimal for $k$, but $c_2, c_3, ..., cm$ is not optimal for $k-c_1$.

There is some other set of coins $c'_2, c'_3, ..., c'_n$ where $n < m$ that add up to $k - c_1$.

$\{c_1, c'_2, c'_3, ..., c'_n\}$ would be a solution, but since $n < m$ this implies that our original solution wasn't optimal!

**Optimal substructure**

Proof by contradiction:

If $\{c_1, c_2, c_3, ..., cm\}$ is optimal for $k$, $\{c_2, c_3, ..., cm\}$ is optimal for $k - c_1$.

We can make greedy decisions.

**Greedy choice property**

Greedy choice property: The greedy choice is contained within some optimal solution.

The greedy choice results in an optimal solution.

Proof by contradiction:

Let $\{c_1, c_2, c_3, ..., cm\}$ be an optimal solution.

Assume it is ordered from largest to smallest.

Assume that it does not make the greedy choice at some coin $c_i$.

Any problem contradiction?
**Greedy choice property**

**Proof by contradiction:**

Let \( \{c_1, c_2, c_3, \ldots, c_m\} \) be an optimal solution

Assume it is ordered from largest to smallest

Assume that it does not make the greedy choice at some coin \( c_i \)

\( g_i > c_i \). We need at least one more lower denomination coin because \( g_i \) can be made up of \( c_i \) and one or more of the other denominations

but that would mean that the solution is longer than the greedy!

---

**Interval scheduling**

Given \( n \) activities \( A = [a_1, a_2, \ldots, a_n] \) where each activity has start time \( s_i \) and a finish time \( f_i \). Schedule as many as possible of these activities such that they don’t conflict.

---

**Which activities conflict?**
Simple recursive solution

Enumerate all possible solutions and find which schedules the most activities

```
INTSCHEDULE-RECURSIVE(A)
1  if A = {} return 0
2  else max = -\infty
3      for all a \in A
4          A' = A minus a and all conflicting actives with a
5          s = INTSCHEDULE-RECURSIVE(A')
6      if a > max
7          max = s
8  return 1 + max
```

Can we do better?

Dynamic programming
- \(O(n^2)\)

Greedy solution – Is there a way to repeatedly make local decisions?
- Key: we'd still like to end up with the optimal solution

Overview of a greedy approach

Greedily pick an activity to schedule

Add that activity to the answer

Remove that activity and all conflicting activities. Call this \(A'\).

Repeat on \(A'\) until \(A'\) is empty
Greedy options

Select the activity that starts the earliest, i.e.
\[ \text{argmin}\{s_1, s_2, s_3, \ldots, s_n\} \]?

Greedy options

Select the activity that starts the earliest, i.e.
\[ \text{argmin}\{s_1, s_2, s_3, \ldots, s_n\} \]?

Greedy options

Select the shortest activity, i.e.
\[ \text{argmin}\{f_1-s_1, f_2-s_2, f_3-s_3, \ldots, f_n-s_n\} \]

Greedy options

Select the shortest activity, i.e.
\[ \text{argmin}\{f_1-s_1, f_2-s_2, f_3-s_3, \ldots, f_n-s_n\} \]

non-optimal

non-optimal
Greedy options

Select the activity with the smallest number of conflicts

Greedy options

Select the activity with the smallest number of conflicts

Greedy options

Select the activity that ends the earliest, i.e. \( \text{argmin}\{f_1, f_2, f_3, \ldots, f_n\} \)
Select the activity that ends the earliest, i.e. \( \text{argmin}\{f_1, f_2, f_3, \ldots, f_n\} \)?

Remove the conflicts
Greedy options

Select the activity that ends the earliest, i.e.
\[ \text{argmin}\{f_1, f_2, f_3, \ldots, f_n\} \]
Greedy options

Select the activity that ends the earliest, i.e. \( \arg\min\{f_1, f_2, f_3, \ldots, f_n\} \)?

Multiple optimal solutions

Greedy options

Select the activity that ends the earliest, i.e. \( \arg\min\{f_1, f_2, f_3, \ldots, f_n\} \)?

Efficient greedy algorithm

Once you've identified a reasonable greedy heuristic:
- Prove that it always gives the correct answer
- Develop an efficient solution
Is our greedy approach correct?

“Stays ahead” argument:

show that no matter what other solution someone provides you, the solution provided by your algorithm always “stays ahead”, in that no other choice could do better.

Let \( r_1, r_2, \ldots, r_k \) be the solution found by our approach

Let \( o_1, o_2, o_3, \ldots, o_k \) be another optimal solution

Show our approach “stays ahead” of any other solution

Stays ahead

Compare first activities of each solution

what do we know?

53

54

Stays ahead

\[ \text{finish}(r_1) \leq \text{finish}(o_1) \]

what does this imply?

55

56
Stays ahead

We have at least as much time as any other solution to schedule the remaining 2...k tasks

An efficient solution

```
INTervALSCHEDULE-GREEDY(A)
1 sort A based on finish times f_i
2 for i ← 1 to n
3 add a_i to R
4 finish ← f_i
5 while s_i < finish
6 i ← i + 1
7 return R
```

Running time?

```
INTervALSCHEDULE-GREEDY(A)
1 sort A based on finish times f_i
2 for i ← 1 to n
3 add a_i to R
4 finish ← f_i
5 while s_i < finish
6 i ← i + 1
7 return R
```

Overall: $\Theta(n \log n)$

Better than:

$O(n!)$

$O(n^2)$

Scheduling all intervals

Given $n$ activities, we need to schedule all activities. Goal: minimize the number of resources required.
Greedy approach?

The best we could ever do is the maximum number of conflicts for any time period

Calculating max conflicts efficiently

Calculating max conflicts efficiently

Calculating max conflicts efficiently

Calculating max conflicts efficiently
Calculating max conflicts efficiently

```
ALL INTERVAL SCHEDULE COUNT(A)
1 Sort the start and end times, call this X
2 current = 0
3 max = 0
4 for i = 1 to length[X]
5 if x_i is a start node
6 current++
7 else current--
8 if current > max
9 max = current
10 return max
```
Correctness?

We can do no better than the max number of conflicts. This exactly counts the max number of conflicts.

\[
\text{ALL Interval Schedule Count(A)}
\]

1. Sort the start and end times, call this \( X \)
2. \( \text{current} \leftarrow 0 \)
3. \( \text{max} \leftarrow 0 \)
4. for \( i \leftarrow 1 \) to \( \text{length}(X) \)
5. if \( x_i \) is a start node
6. \( \text{current} \leftarrow + \)
7. else
8. \( \text{current} \leftarrow - \)
9. if \( \text{current} > \text{max} \)
10. \( \text{max} \leftarrow \text{current} \)
11. return \( \text{max} \)

Runtime?

\[
O(2^n \log 2n + n) = O(n \log n)
\]

\[
\text{ALL Interval Schedule Count(A)}
\]

1. Sort the start and end times, call this \( X \)
2. \( \text{current} \leftarrow 0 \)
3. \( \text{max} \leftarrow 0 \)
4. for \( i \leftarrow 1 \) to \( \text{length}(X) \)
5. if \( x_i \) is a start node
6. \( \text{current} \leftarrow + \)
7. else
8. \( \text{current} \leftarrow - \)
9. if \( \text{current} > \text{max} \)
10. \( \text{max} \leftarrow \text{current} \)
11. return \( \text{max} \)

Knapsack problems:

Greedy or not?

0-1 Knapsack – A thief robbing a store finds \( n \) items worth \( v_1, v_2, \ldots, v_n \) dollars and weight \( w_1, w_2, \ldots, w_n \) pounds, where \( v \) and \( w \) are integers. The thief can carry at most \( W \) pounds in the knapsack. Which items should the thief take if he wants to maximize value.

Fractional knapsack problem – Same as above, but the thief happens to be at the bulk section of the store and can carry fractional portions of the items. For example, the thief could take 20% of item \( i \) for a weight of 0.2\( w_i \) and a value of 0.2\( v_i \).