Analysis of Algorithms
Homework 1 – Greedy Algorithms

Abraham Murciano

December 28, 2020

1   The 0-1 Knapsack Problem

We are given a set of n items and a knapsack which can carry up to a weight of W. For each item i, its weight is w_i and its value v_i. Each item can be put into the knapsack only as a whole. The problem is to find a subset of items of maximal value which can be carried in the knapsack without exceeding the weight limit.

Part A

Suppose there is a greedy algorithm which attempts to solve this problem, defined as follows.

---
function GREEDYKNAPSACK(items, w, v, W)
    knapsack := {}
    current weight := 0
    current value := 0
    sort items by value to weight ratio (highest first)
    for i in items **do**
        if current weight + $w_i \leq W$ **then**
            insert i into knapsack
            increment current weight by w_i
            increment current value by v_i
    return knapsack

Now suppose we attempt to use this algorithm to give us the optimal knapsack for a list of items with the following parameters.

$$\text{items} = \{1, 2, 3\}$$
$$W = 9$$
$$w_1 = 6, \ w_2 = 5, \ w_3 = 4$$
$$v_1 = 7, \ v_2 = 5, \ v_3 = 4$$

Our algorithm would first sort the items by value to weight ratio, and since item 1 has a slightly higher ratio it will be the first item. Then the algorithm will check if item 1 fits into the knapsack, which it does. Therefore it will be inserted. Now, when it moves onto the next two items, neither will fit, since the knapsack already has item one. So the knapsack returned will only contain item 1, thus yielding a total value of 7.

However, this is not the optimal solution, since we can fit both items 2 and 3, for a total weight of 9, and a total value of 9.

Part B

The example described in Part A satisfies the condition that when the weights are sorted in increasing order, namely $(4, 5, 6)$, the values also turn out to be sorted in increasing order, namely $(4, 5, 7)$. As shown in Part A, this is in fact a counter example to the correctness of a greedy algorithm, so we may conclude that the greedy algorithm defined above is not necessarily correct even when the weights and values are in the same order.

2 Coin Exchange Problem

We are given a banknote of value A and (an unlimited amount of) coins of values $\{c^i : c, i \in \mathbb{N} \land 0 \le i \le k \land c > 1 \land k > 1\}$. The goal is to find a way to change the banknote for the minimal number of coins.

Part A

Suppose we have a problem of this format, whose optimal solution to reach the value A is some set of coins x (possibly including many different coins of the same value), we can be certain that for any subset of these coins $s \subseteq x$, s will be an optimal solution to reach the value which the coins in s sums up to.

This can be easily proven since if this was not the case, and there was a set t such that $|t| < |s|$, whose coins summed up to the same values as those in s, then we would have a solution to the initial problem $y = (x - s) \cup t$ such that $|y| < |x|$. And this is a contradiction to our assumption that x was an optimal solution.

Part B

A greedy algorithm that solves this problem is as follows.

---
function GREEDYCOINEXCHANGE(c, k, A)
    if $A = 0$ **then return** ϕ
    $a := \text{MIN}(\lfloor log_c A \rfloor, k)$
    return $\{$a coin of value $c^a\} \cup$ GREEDYCOINEXCHANGE$(c, k, A - c^a)$

To prove the correctness of this algorithm, we must show that the solution obtained by it is always optimal. We will do this by use of a couple of lemmas.

Lemma 1. The first lemma we must use is as follows. Suppose that we have a set of coins S (possibly with multiple coins of the same denomination) whose values are all of the form c^w, where $w < k$, c is a constant, and w may change from coin to coin. Additionally, suppose that $\sum_{s \in S}$ value$(s) \geq c^k$. Then there exists some set $T \subseteq S$ such that $\sum_{t \in T}$ value$(t) = c^k$.

To prove this lemma, let $T \subseteq S$ be the subset with the smallest sum greater than or equal to c^k. T must exist, since in absence of any other choices, S would fit the criteria. So we suppose our claim is false, and the sum of the coins in T is V so

$$V = \sum_{t \in T} \text{value}(t) > c^k$$

Now let c^z be the value of the smallest coin in T. Since all the coins in T have values of sums of higher powers of c, V must be divisible by c^z. And since c^k is also divisible by c^z (because $z < k$), therefore $V - c^k$ must be divisible by c^z, so

$$V - c^k \geq c^z \quad \Rightarrow \quad V - c^z \geq c^k$$

If this was true, we could remove a coin of value c^z from T to obtain a subset of S with sum still greater than or equal to c^k, contradicting our assumption that T was the subset with the smallest sum greater than or equal to c^k. Therefore our claim is true and there does exist a $T \subseteq S$ such that $\sum_{t \in T}$ value$(t) = c^k$.

Lemma 2. Our second lemma states that the optimal solution to reach the value A must contain at least one coin of the highest available value less than or equal to A.

We will prove the lemma by contradiction. Suppose the optimal solution to reach a goal of A did not contain the largest available coin value less than or

equal to A. We'll call this value c^a, as labelled in the algorithm. That means that the optimal solution is a set of coins S, all smaller than or equal to c^{a-1}, which reach A, and reach or exceed c^a. However, by our first lemma, there must be some subset of S which exactly equals c^a, since all coins in S are powers of c. Therefore, we would be able to replace said subset with a single coin of value c^a, thus giving us a more optimal solution than S, showing that S was not optimal, which is a contradiction. Thus our lemma must be correct.

The Proof. We can now prove the correctness of the algorithm using induction. With a base case of $A = 1$, the algorithm will return a set containing a single coin of value 1, which is trivially the optimal solution.

Now we assume that for all $A < u$, GREEDYCOINEXCHANGE(c, k, A) returns an optimal solution, and show from there that GREEDYCOINEXCHANGE(c, k, u) also does. Our algorithm returns a set containing the largest coin available to us which does not exceed u, as well as the optimal solution to the remaining value. And since according to our lemma the optimal solution must contain said coin, our solution must be optimal.

Part C

The runtime of the algorithm is $O(\lg(A))$ if the largest coin has a value larger than A, or $O(A)$ if A is much larger than all the coins. All things considered, the complexity is $O(A - k + \lg(A))$.

Part D

There is no greedy solution for a problem of this sort, where the coin denominations are arbitrary, and are not powers of some common base. As a counter-example, suppose we have the coin denominations $\{10, 9, 1\}$. If we apply a greedy algorithm to reach a total value of eighteen, the first coin we would take would have a value of ten, and then we would have to take eight more coins with a value of one. The solution would contain a total of nine coins. However, the optimal solution is two coins worth nine each.

3 Relay Stations Problem

There is a straight stretch of road on which a set of n houses H are arbitrarily scattered. We are given that the distance from the beginning of the road to the ith house h_i is d_{h_i}. The goal is to position a minimal number of relay stations such that each house is within some range R from at least one relay station.

Part A

Suppose p is a minimal set containing the distances to place relay stations to cover all the houses. Then we can say that for any $r \subseteq p$, it will be a minimal solution to cover all the houses which r covers (with a distance no larger than R).

This can be proven by contradiction. Suppose the contrary; that there exists a set of station positions s such that $|s| < |r|$ and placing a station in every position in s will cover all the houses which r covers. Then we can obtain another solution $q = (p - r) \cup s$ where $|q| < |p|$, contradicting the minimality of p. Therefore r must be a minimal solution to cover the houses which it covers.

Part B

---
function RELAYSTATIONS(H, R)
    sort H
    $f := -\infty$ $\triangleright$ f is the distance to the first unresolved house
    $p := \phi$ $\triangleright$ p is the positions we will place relay stations
    for $h \in H$ **do**
        if $d_h > f + 2R$ **then** $\triangleright$ h no longer fits into previous group
            $p := p \cup \{f + R\}$ $\triangleright$ Position a station in for the previous group
              $f := d_h$ $\triangleright$ Start a new group at the next house
    return p

Proof of Correctness. Suppose the set of station positions returned by our algorithm is p. Suppose q is a minimal solution. None of the relay stations positioned by our algorithm are within the distance $2R$ of another, that is, their signals do not overlap.

Define p_i as the ith smallest position to the beginning of the road in p, and define q_i similarly. It must be the case that $\forall i (p_i \geq q_i)$, meaning that the ith station according to our algorithm is not closer to the start than the ith station in any minimal solution. This is because if for some i, q_i was further down the road than p_i, at least one house (the one furthest up the road which is covered by p_i) would not have coverage from q's placements, which is impossible since q must cover all the houses.

It is fairly obvious that both the furthest station in p and the furthest station in q give service to the furthest house. So the distance between them is at most $2R$.

Now to conclude our proof, we suppose that p is not minimal. That means $|q| = a < |p| = b$. And we have shown that p_a must be further down the road than q_a, which is the position of last station in q. However, p_b (the position

of the last station in p), which is greater than p_a, must be within $2R$ of q_a. Meaning $q_a \leq p_a < p_b$, which implies $p_b - p_a \leq p_b - q_a \leq 2R$. Which means that p_a and p_b have overlapping ranges, which is a contradiction. Therefore $|p| = |q|$, so p is minimal.

Part C

Quite clearly, the complexity of the second part of RELAYSTATIONS is linear with respect to the number of houses. That is, $O(n)$. However, we must first sort the houses by their distance from the start of the road, so overall, the complexity is $O(n + n \lg(n)) = O(n \lg(n))$.