CS 1510: Algorithm Design
Homework 4 - Greedy Problem 9
Zach Sadler, John Hofrichter
zps6@pitt.edu — jmh162@pitt.edu
September 5, 2013

Problem 9a

We will disprove that this algorithm is correct. Consider the input

$$\text{Skier heights} = \{5, 13\}$$
$$\text{Ski heights} = \{12, 20\}$$

When given this input, the algorithm will first compute the four combinations of ski and skier and find the minimum difference, specifically the algorithm will select the Skier with height 13 and the Ski with height 12, whose difference is $|13 - 12| = 1$. After this choice, the algorithm is forced to select the remaining pair, the Skier with height 5 and the Ski with height 20, whose difference is $|5 - 20| = 15$. Thus the average difference is

$$\frac{15 + 1}{2} = 8$$

However, this is not the optimal output. A better output is to pair Skier with height 5 and Ski with height 12 whose difference is $|5 - 12| = 7$, then pair Skier with height 13 and Ski with height 20 whose difference is $|13 - 20| = 7$. Thus the average difference for these two pairs is

$$\frac{7 + 7}{2} = 7$$

Since $7 < 8$, we found an acceptable output which is less than the algorithm's output, and thus we have a counterexample which disproves the proposition that the algorithm is correct.

Problem 9b

We will prove that the algorithm is correct.
Let G be the greedy algorithm which pairs the shortest Skier with the shortest Ski, the second shortest Skier with the second shortest Ski, and so on. Suppose towards a contradiction that there is an input I on which G gives

incorrect output $G(I)$. Then by definition, $G(I)$ does not produce a series of pairs of Skis and Skiers whose average difference between heights of pairs is minimized.

Note that the output of both algorithms is a sequence of pairs of a Skier and his Ski. We will arrange this output in order of Skiers from smallest to tallest. Now we will define $Opt(I)$ to be the algorithm which produces the minimum average difference between pairs of Skier and Ski (the correct answer) and agrees with $G(I)$ for the maximum number of steps, $k-1$. We will also specify that $Opt(I)$'s output is also arranged in order of increasing Skier height. This is well defined, since we are ordering our output from smallest Skier to tallest, so $Opt(I)$ agrees with $G(I)$ for $k-1$ Skiers consecutively from the beginning, which is at least as many steps as any other optimal solution.

Now define P_k (person k) to be the kth shortest Skier, and S_k his associated Ski as given in $G(I)$. We know that $G(I)$ and $Opt(I)$ have agreed up to this point, and we also know that $Opt(I)$ selects a different Ski for P_k, which we will name S_a. Since Opt did not pick Ski S_k for P_k, we know that Opt must use S_k with some other Skier, which we will name P_b.

Now we will define $Opt'(I)$ by copying $Opt(I)$ but pairing P_k with S_k as in G(I), and pairing P_b with S_a. Thus $Opt'(I)$ agrees with $G(I)$ for one more term (the $\{P_k, S_k\}$ term) than $Opt(I)$.

Next we must show that $Opt'(I)$ is still optimal.

First, note that

$$P_k \leq P_b$$

since we have arranged our output in increasing order of Skier heights, and P_b has not occurred before P_k since $G(I)$ and $Opt(I)$ agreed up to P_k and so P_b certainly occurs later in $Opt(I)$ than P_k.

Next, note that

$$S_k \leq S_a$$

since by definiton G pairs the shortest Ski with the current (shortest) Skier, and so since S_a is paired with a Skier of greater or equal height than S_k, we know that S_k itself is not greater than S_a.

Combining these two inequalities, we are left with 6 possibilities:

$$P_k \leq P_b \leq S_k \leq S_a \tag{1}$$
$$P_k \leq S_k \leq P_b \leq S_a \tag{2}$$
$$P_k \leq S_k \leq S_a \leq P_b \tag{3}$$
$$S_k \leq S_a \leq P_k \leq P_b \tag{4}$$
$$S_k \leq P_k \leq S_a \leq P_b \tag{5}$$
$$S_k \leq P_k \leq P_b \leq S_a \tag{6}$$

We want to show that if $Opt(I)$ is optimal then $Opt'(I)$ is also optimal, so we want to show that our single swap of S_k and S_a in $Opt(I)$ will not increase the average difference between Skier and Ski. That is, we want to show that

$$|P_k - S_k| + |P_b - S_a| \leq |P_k - S_a| + |P_b - S_k|$$

We will show that this inequality holds for all 6 cases by 'breaking' the absolute values according to which of the 2 terms in the subtraction statement are larger in the given case. You can see this as equivalent to beginning with the 6 inequality cases and adding or subtracting equivalent terms to each side to acheive the above inequality.
Case 1:

$$(S_k - P_k) + (S_a - P_b) \leq (S_a - P_k) + (S_k - P_b)$$
$$0 \leq 0$$

which is true, since $0 = 0$.

Case 2:

$$(S_k - P_k) + (S_a - P_b) \leq (S_a - P_k) + (P_b - S_k)$$
$$S_k - P_b \leq P_b - S_k$$
$$S_k \leq P_b$$

which is true by assumption.

Case 3:

$$(S_k - P_k) + (P_b - S_a) \leq (S_a - P_k) + (P_b - S_k)$$
$$S_k - S_a \leq S_a - S_k$$
$$S_k \leq S_a$$

which is true by assumption.

Case 4:

$$(P_k - S_k) + (P_b - S_a) \leq (P_k - S_a) + (P_b - S_k)$$
$$0 \leq 0$$

which is true, since $0 = 0$.

Case 5:

$$(P_k - S_k) + (P_b - S_a) \leq (S_a - P_k) + (P_b - S_k)$$
$$P_k - P_b \leq P_b - P_k$$
$$P_k \leq P_b$$

which is true by assumption.

Case 6:

$$(P_k - S_k) + (S_a - P_b) \leq (S_a - P_k) + (P_b - S_k)$$
$$P_k - P_b \leq P_b - P_k$$
$$P_k \leq P_b$$

which is true by assumption.

Thus we have shown that in all 6 possible cases for the ordering of P_k, P_b, S_k, S_a, we have that

$$|P_k - S_k| + |P_b - S_a| \leq |P_k - S_a| + |P_b - S_k|$$

holds true. That is, Opt' is optimum if Opt is.
Thus we have constructed $Opt'(I)$ which is optimum and agrees with $G(I)$ for one more step than $Opt(I)$, and so we have a contradiction that $Opt(I)$ agrees with $G(I)$ more than any other optimum output, and we are done.

Unrelated: Was this too verbose? Not enough? Can you give me some feedback on if I need to be so explicit and talk through the process, or if I can be a little more loose with my proofs? I just don't want to give you (the grader/reader) more information than you need to the point where I'm repeating myself and being unclear. Thanks.