0.0.1. The Turnpike Reconstruction Problem

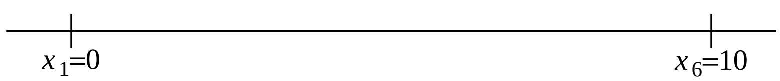
Suppose we are given n points $p_1, p_2, \ldots, p_n$ located on the $x-$ axis. x_i is the $x-$coordinate of p_i. Let us further assume that $x_1=0$, and the points are given from left to right. These n points determine $n(n-1)/2$ (not-necessarily unique) distances $d_1, d_2, \ldots d_n$ between every pair of points of the form $|x_i - x_j|$ $(i \neq j)$. It is clear that if we are given the set of points, then it is easy to *construct* the set of distances. The *Turnpike reconstruction problem* is to reconstruct a point set from the distances. This finds applications in physics and molecular biology (see the references for pointers to more specific information). The name derives from the analogy of points to turnpike exits on east-coast highways. Just as factoring seems harder than multiplication, the reconstruction problem seems harder than the construction problem. Nobody has been able to give an algorithm which is guaranteed to work in polynomial time. The algorithm which we will present seems to run in $O(n^2 \log n)$; no counter-example to this conjecture is known, but it is still just that -- a conjecture.

Of course given one solution to the problem, an infinite number of others can be constructed by adding an offset to all the points. This is why we insist that the first point is anchored at zero and that the points set which constitutes a solution is output in non-decreasing order.

Let D be the set of distances, and assume that $|D| = m = n(n-1)/2$. As an example, suppose that

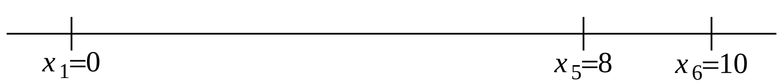
D = { 1, 2, 2, 2, 3, 3, 3, 4, 5, 5, 5, 6, 7, 8, 10 }.

Since $|D|=15$, we know that $n=6$. We start the algorithm by setting $x_1=0$. Clearly, $x_6=10$, since 10 is the largest element in D. We remove 10 from D. The points which we have placed, and the remaining distances are:



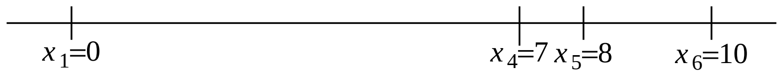
D = { 1, 2, 2, 2, 3, 3, 3, 4, 5, 5, 5, 6, 7, 8 }.

The largest remaining distance is 8, which means that either $x_2=2$ or $x_5=8$. By symmetry, we can conclude that the choice is unimportant since either both choices lead to a solution (which are mirror-images of each other), or neither do, so we can set $x_5=8$, without affecting the solution. We then remove the distances $x_6 - x_5 = 2$ and $x_5 - x_1 = 8$ from D, obtaining



D = { 1, 2, 2, 3, 3, 3, 4, 5, 5, 5, 6, 7 }.

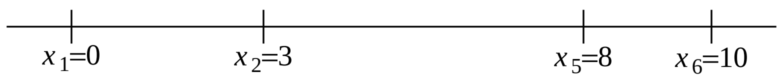
The next step is not obvious. Since 7 is the largest value in D, either $x_4=7$ or $x_2=3$. If $x_4=7$, then the distances $x_6-7=3$, $x_5-7=1$, must also be present in D. A quick check shows that indeed they are. On the other hand, if we set $x_2=3$, then $3-x_1=3$ and $x_5-3=5$ must be present in D. These distances are also in D, so we have no guidance on which choice to make. Thus we try one, and see if it leads to a solution. If it turns out that it doesn't, we can come back and try the other. Trying the first choice, we set $x_4=7$, which leaves



D = { 2, 2, 3, 3, 4, 5, 5, 5, 6 }.

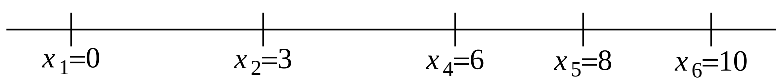
At this point, we have x_1=0, x_4=7, x_5=8, and x_6=10. Now the largest distance is 6, so either x_3=6 or x_2=4. But if x_3=6, then x_4-x_3=1, which is impossible since 1 is not in D. On the other hand, if x_2=4, then x_2-x_0=4, and x_5-x_2=4. This is also impossible, since 4 only appears once in D. Thus, this line of reasoning leaves no solution, so we backtrack.

Since x_4=7 failed to produce a solution, we try x_2=3. If this also fails, then we give up and report no solution. We now have



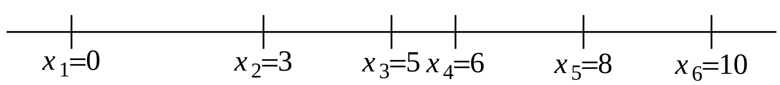
D = { 1, 2, 2, 3, 3, 4, 5, 5, 6 }.

Once again, we have to choose between x_4=6 and x_3=4. x_3=4 is impossible because D only has one occurrence of 4, and two would be implied by this choice. x_4=6 is possible, so we obtain



D = { 1, 2, 3, 5, 5}.

The only remaining choice is to assign x_3=5; this works because it leaves D empty, and so we have a solution.



D = { }.

Fig. 0.1 shows a decision tree representing the actions taken to arrive at the solution. Instead of labelling the branches, we've placed the label in the branches' destination node. A node with an asterisk indicates that the points chosen are inconsistent with the given distances; nodes with two asterisks have only impossible nodes as children, and thus represent an incorrect path.

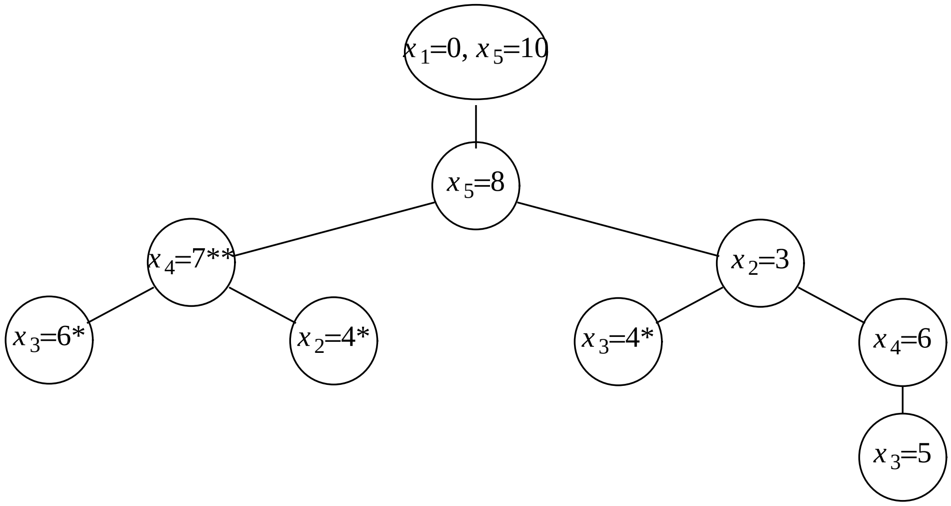


Fig. 0.1. Decision tree for the worked Turnpike Reconstruction example.