

Problem Set 1

Due: Monday, September 22nd, 2014

Problem 1. For each of the following problems, either show that the problem is in P by giving a polynomial-time algorithm (e.g., by reducing to shortest paths, network flow, matching, or minimum spanning tree); or show that the problem is NP-hard by reducing from 3-Partition, 3-Dimensional Matching, or Numerical 3-Dimensional Matching.

- (a) Given a multiset of non-negative integers $A = \{a_1, \dots, a_{2n}\}$ that sum to tn , find a partition of A into n groups S_1, \dots, S_n of size 2 such that each group sums to t .

Solution: Create a graph with one vertex for each input number. For each pair of numbers a_i, a_j check whether $a_i + a_j = t$. If so, add an edge to the graph. Otherwise, there should be no edge between a_i and a_j . Hence, each edge represents a possible group in the partition. Next, run a matching algorithm. If the result is a perfect matching, we can construct the corresponding partition by creating one group for each edge in the matching. Because we start with a matching, each a_i can belong to at most one group in the corresponding partition. And because the matching that we start with is perfect, we are guaranteed to have n groups of size 2.

The converse is also true. Suppose that we have a partition satisfying the problem constraints. For each group $\{a_i, a_j\}$ in the partition, we are guaranteed that $a_i + a_j = t$, so the corresponding edge (a_i, a_j) must exist. Hence, we can add it to the matching. Because we started with a partition, no two edges in the matching share an endpoint. And because the number of groups is n while the number of vertices is $2n$, we know that the matching constructed in this fashion must be perfect. \square

Alternative Solution: Assume without loss of generality that the numbers a_1, \dots, a_{2n} are in sorted order. Suppose, for the sake of contradiction, that there is at least one solution to the problem, but that no solution contains the group $\{a_1, a_{2n}\}$. Pick an arbitrary solution, and let a_i, a_j be such that $\{a_1, a_i\}$ and $\{a_j, a_{2n}\}$ are sets in the solution we picked. Then $a_i = t - a_1$, and $a_j = t - a_{2n}$. Because the numbers are in sorted order, we know that $a_1 \leq a_i, a_j \leq a_{2n}$. Hence, we have $a_1 \leq t - a_{2n}$, and $t - a_1 \leq a_{2n}$. A little algebra shows that this is equivalent to saying that $t \leq a_1 + a_{2n} \leq t$, and therefore $a_1 + a_{2n} = t$. This means that $a_i = a_{2n}$, and $a_j = a_1$. Therefore, we can modify our solution to remove the groups $\{a_1, a_i\}$, $\{a_j, a_{2n}\}$, and replace them with groups $\{a_1, a_{2n}\}$ and $\{a_i, a_j\}$ without affecting the correctness. But this contradicts our assumption that there is no solution that contains the group $\{a_1, a_{2n}\}$. Therefore, we know that if there is any solution at all, there must be some solution containing $\{a_1, a_{2n}\}$. Hence, we may safely pair those two elements, and recurse on the remaining elements, so that a_i is always paired with a_{2n+1-i} . If this yields groups that sum to t , we have a solution. Otherwise, we know that there can be no solution. \square

- (b) Given a multiset of non-negative integers $A = \{a_1, \dots, a_{2n}\}$ that sum to tn , find a partition of A into n groups S_1, \dots, S_n of any size such that each group sums to t .

Solution: Reduce from standard 3-Partition (the variant where any number of numbers is allowed to belong to a single group). Let a_1, \dots, a_{3n} be the groups of input numbers. Define a new sequence of numbers b_1, \dots, b_{4n} as follows:

$$b_i = \begin{cases} a_i & \text{if } i \leq 3n \\ t & \text{otherwise} \end{cases}$$

Suppose that we are given a partition of these numbers into $2n$ groups that sum to t . Clearly, $b_{3n+1}, \dots, b_{4n} = t$, so any group containing one of those n numbers cannot contain any other numbers. Hence, the remaining n groups must contain all $3n$ numbers from the original 3-Partition instance. Furthermore, each group must sum to t . Hence, the assignment of those numbers to the remaining n groups tells us how to solve the original 3-Partition problem. \square

Alternative Solution: For the variant originally posted, where the input numbers form a set (and thus do not contain any duplicates), we may show that the problem is in P as follows. Because we have $2n$ numbers divided among n groups, the average number of numbers per group is equal to 2. An empty group sums to 0, while t must be strictly greater than 0. There may be a single group of size 1, with one number equal to t , but because the input numbers form a set, there cannot be any other groups with one element that sum to t .

Hence, we may solve the problem as follows. If A does not contain t , then the minimum size of each group is 2, and therefore each group must have size 2, reducing us to the situation from part (a). Otherwise, there must be one group with one element (containing t), one group with three elements (since the average number of elements per group is 2, and only one group contains fewer than 2 elements), and $n - 2$ groups with 2 elements. For each triple $T \subset S$ with size $|T| = 3$, try forming a group of three with T , testing whether the sum of T is t , and then solving the rest of the problem ($A \setminus (T \cup \{t\})$) using the algorithm from part (a). If we succeed with any of these calls, then we have a solution to the overall problem. And because we have shown that any solution where $t \in A$ must have this form (one group of 1, one group of 3, and $n - 2$ groups of 2), we know that if there is no solution of this form, then there is no solution overall. \square

- (c) Given a multiset of non-negative integers $A = \{a_1, \dots, a_{2n}\}$ and a sequence of target numbers $\langle t_1, \dots, t_n \rangle$, find a partition of A into n groups S_1, \dots, S_n of size 2 such that for each $i \in \{1, \dots, n\}$, the sum of the elements in S_i is t_i .

Solution: Reduction from Numerical 3-Dimensional Matching. Given $A = \{a_1, \dots, a_n\}$, $B = \{b_1, \dots, b_n\}$, and $C = \{c_1, \dots, c_n\}$, with target sum t , we define the new numbers as follows:

$$d_{2(i-1)+1} = 2a_i + 0$$

$$d_{2(i-1)+2} = 2b_i + 1$$

And the target values are:

$$q_i = 2(t - c_i) + 1$$

Suppose that we have a sequence of groups S_1, \dots, S_n satisfying the desired constraints. By examining the targets modulo 2, we can see that each group must contain exactly one

number that is equivalent to 1 mod 2. By construction, only the numbers $d_{2(i-1)+2} \equiv 1 \pmod{2}$. Therefore, each group must contain exactly one number $d_{2(i-1)+2} = 2b_i + 1$, and the other number in each group must be some $d_{2(j-1)+1} = 2a_j + 0$. So for the k th group, we must have:

$$\begin{aligned} d_{2(i-1)+2} + d_{2(j-1)+1} &= q_k \\ 2b_i + 1 + 2a_j + 0 &= 2(t - c_k) + 1 \\ a_j + b_i + c_k &= t \end{aligned}$$

Which is precisely what we wanted.

Conversely, suppose that we have a solution to the original Numerical 3-Dimensional Matching instance. Then for each group $\{a_i, b_j, c_k\}$, we set $S_k = \{d_{2(i-1)+1}, d_{2(j-1)+2}\}$. We are guaranteed that $a_i + b_j + c_k = t$, so we have:

$$\sum S_k = 2a_i + 0 + 2b_j + 1 = 2(a_i + b_j) + 1 = 2(t - c_k) + 1 = q_k$$

□

Problem 2. Give a direct reduction from 3-Partition to Partition. (*Hint:* First reduce directly from 3-Partition to Subset-Sum, then modify the proof to work with Partition.)

Solution: Let a_1, \dots, a_{3n} be the multiset of numbers to partition, and let T be the target sum for each group. For each number a_i and each possible group $k \in \{0, \dots, n-1\}$, we add the following number to our 2-Partition instance:

$$x_{i,k} = 1 \cdot (Tn)^{n+i} + a_i \cdot (Tn)^k$$

The target number we would aim for in a Subset-Sum problem would be:

$$T' = \sum_{i=1}^{3n} 1 \cdot (Tn)^{n+i} + \sum_{k=0}^{n-1} T \cdot (Tn)^k$$

Consider the values mod (Tn) . Clearly, $T \pmod{(Tn)} = T$, and for $k \neq 0$, $x_{i,k} \pmod{(Tn)} = 0$. So in order to get our target sum, we need to use a subset of the numbers $x_{1,0}, \dots, x_{n,0}$ that sums to $T \pmod{(Tn)}$. By construction, this is equivalent to finding a subset of the numbers a_1, \dots, a_{3n} that sums to T , and then using the corresponding numbers $x_{i,0}$ in our Subset-Sum problem. A similar argument shows that, for any $k \in \{0, \dots, n-1\}$, we must pick numbers $x_{i_1,k}, \dots, x_{i_q,k}$ such that $a_{i_1} + \dots + a_{i_q} = T$. Furthermore, if we examine the sum mod $(Tn)^{n+i+1}$ for each $i \in \{1, \dots, 3n\}$, it is clear to see that for each number $i \in \{1, \dots, 3n\}$, we can pick only one $x_{i,k}$ to belong to our subset sum. Hence, if we can find a subset of numbers that sums to the target, we know that there must exist a partition of a_1, \dots, a_{3n} into n groups, each of which sums to T .

Next, we wish to convert our reduction to Subset-Sum into a reduction to Partition. The sum of all numbers in our problem is

$$\begin{aligned} Q &= \sum_{i=1}^n \sum_{k=0}^{n-1} x_{i,k} \\ &= \sum_{i=1}^n \sum_{k=0}^{n-1} \left(1 \cdot (Tn)^{n+i} + a_i \cdot (Tn)^k \right) \\ &= \sum_{i=1}^n n \cdot (Tn)^{n+i} + \sum_{k=0}^{n-1} (Tn) \cdot (Tn)^k \end{aligned}$$

To ensure that we find a subset that sums to T , we add one extra number $Q - 2T$. (Note that because Q is very large in comparison to T , this new number will not be negative.) With this extra number, the new total becomes $2Q - 2T$, so a solution to the Partition problem must make both halves sum to $Q - T$. One of those halves must contain the extra number $Q - 2T$, so the set of all other numbers in that half must sum to $(Q - T) - (Q - 2T) = T$, which is precisely what we wanted. \square

Problem 3. Suppose you are given a weighted connected undirected graph $G = (V, E, w)$ satisfying the triangle inequality—that is, for any three vertices $x, y, z \in V$ connected in a triangle $(x, y), (y, z), (x, z) \in E$, we have $w(x, z) \leq w(x, y) + w(y, z)$. Your goal is to assign each node one of k colors. Define the *total weight* of a color be the sum of all of the distances between pairs of nodes of that color; where distance is the weight of the minimum weight path between the nodes. Show that it is NP-complete to find a color assignment in which the total weight of each color is less than t .

Solution: The problem is obviously in NP with a possible certificate being the vertex assignment. We will prove this problem is NP-Complete by reducing from 3-Partition. For every number p_i in our 3-Partition instance we will assign a vertex v_i . We now construct a complete graph over these vertices with the weight of the edge between v_i and v_j equal to $(p_i + p_j)/2$. This weighting obeys the triangle inequality. If we pick some third vertex v_l then the weight from v_i to v_j to v_l is equal to $(p_i + p_l)/2 + (p_l + p_j)/2 = (p_i + p_j)/2 + p_l \geq (p_i + p_j)/2$. We let the number of colors be equal to $n/3$ where n is the number of elements in the 3-Partition. Set $t = \frac{3}{n} \sum_{i=1}^n p_i$. Notice, if we assign each color to exactly three vertices, v_i, v_j, v_k , the total weight of that color will be equal to $p_i + p_j + p_k$. Now, if any color has fewer than 3 vertices, at least one color must have more than three vertices by the Pigeon Hole Principal. If this is the case, the total weight is equal to 1.5 times the sum of four elements from our 3-Partition. Since we are guaranteed these are between $t/4$ and $t/2$, any coloring of this sort cannot be a solution. Thus if there is a solution each color must have three vertices which corresponds to a solution of our 3-Partition, and there is only a solution if there is a solution to the corresponding 3-Partition problem. \square

Problem 4. For each of the following problems, either show that it can be solved in polynomial time, or prove that the problem is NP-hard.

- (a) You are trying to solve a $\sqrt{n} \times \sqrt{n}$ (unsigned) square edge-matching puzzle, which originally had n pieces. Unfortunately, you've managed to misplace $2/3$ of the puzzle pieces, leaving you with only $n/3$ pieces. A *configuration* of such a “partial” puzzle is a mapping of the remaining pieces onto the original $\sqrt{n} \times \sqrt{n}$ lattice; a configuration is *valid* if any two remaining pieces mapped to adjacent places match at their touching edges. How hard is it to solve (find a valid configuration of) the puzzle now?

Solution: Our goal is to fit $n/3$ pieces of the edge-matching puzzle onto a board of size $\sqrt{n} \times \sqrt{n}$ such that there are no edge mismatches. By partitioning the potential positions into two sets in a checkerboard fashion and selecting the larger half, we can construct a set of at least $n/2$ cells such that no two cells are adjacent. In those $n/2$ cells, we can place any pieces we desire — in particular, we can place the $n/3$ pieces that we do have. Hence, it is always possible to solve the puzzle. \square

- (b) Several weeks later, while digging through the attic, you unearth another $1/3$ of the puzzle pieces, bringing you up to a total of $2n/3$ pieces of the original $\sqrt{n} \times \sqrt{n}$ puzzle. How hard is it to solve the puzzle now?

Solution: Reduction from 3-Partition, with structure borrowed from the original edge-matching reduction. Let $A = \{a_1, \dots, a_{3n}\}$ be the 3-Partition instance, with $t = (\sum A)/n$ the target sum for each group. Without loss of generality, assume that $t \geq n$. Set $k = 2(t+2)$, and let $m = k^2$ be the size of the original puzzle. We start by filling the puzzle with pieces, such that each color on the boundary is unique, and each color on the interior is used only by two pieces. We then cut out a box of width t and height n from the upper left corner, offset by 1 from the top and left. Call this the “frame.” On the sides of this box, we use the color red; on the top of the box, we use the color green. We then complete the reduction much like the original edge-matching reduction: for each a_i , we create a bar of tiles of width a_i , with red edge on either end, and green edges on the top and bottom. The internal edges will have color c_i .

To reduce the number of pieces, we examine the set of pieces on the strict interior of the frame. Those pieces can be partitioned into two sets that form a checkerboard (so that no pair of pieces from the same set are adjacent). Furthermore, one of those sets must contain more than half of the pieces on the strict interior of frame. We then remove a subset of the larger half of size $m/3$.

First, we wish to verify that the larger half has at least $m/3$ puzzle pieces to remove. The frame contains $k^2 - tn$ pieces in total; the boundary of the frame contains $4(k-1)$ pieces on the boundary of the square, and another $t+n-1$ pieces on the boundary of the inset box. Therefore, the set of pieces we may remove has size at least

$$\begin{aligned} & (k^2 - tn - 4(k-1) - (t+n-1))/2 \\ &= (4(t+2)^2 - tn - (8t+12) - (t+n-1))/2 \\ &\geq (4t^2 + 16t + 16 - t^2 - 8t - 12 - t - t + 1)/2 \\ &= (3t^2 + 6t + 5)/2. \end{aligned}$$

By construction, $m = k^2 = 4t^2 + 16t + 16$, so we are guaranteed to have enough pieces as long as:

$$\begin{aligned} m/3 &\leq (3t^2 + 6t + 5)/2 \\ (4t^2 + 16t + 16)/3 &\leq (3t^2 + 6t + 5)/2 \\ 8t^2 + 32t + 32 &\leq 9t^2 + 18t + 15 \\ 14t + 17 &\leq t^2 \end{aligned}$$

So as long as $t \geq 16$, we can remove the correct number of pieces.

Next, we wish to show that this puzzle can be solved if and only if there is a solution to the 3-Partition instance. By construction, if we have a 3-Partition solution, we can construct the corresponding puzzle solution by putting the frame together the way we constructed it, and then slotting each width- a_i bar into the row corresponding to the section of the partition. This works for the same reason as the original edge-matching reduction.

Suppose instead that we have a solution to the puzzle. Let $q = m/3$ be the number of pieces we removed. Consider the number of unique colors. Because each piece we removed is adjacent on all four sides to pieces that were not removed, we know that the number of unique colors in the remaining puzzle pieces is $4k + 4q$: $4k$ from the four sides of the puzzle, and $4q$ from the four sides of the q pieces we removed. And because there are only q pieces missing from our puzzle, we know that there are at most $4k + 4q$ edges that can be adjacent to empty squares. Hence, for each color in our edge-matching puzzle, if an edge is adjacent to an empty square, then it must be one of the $4k + 4q$ unique colors. And if an edge doesn't have a unique color (that is, if it can be matched), it must either be on the boundary, or adjacent to another puzzle piece.

In particular, consider the set of puzzle pieces on the boundary of the frame. Each color between two pieces in the frame boundary occurs exactly twice, once for each piece, and is therefore matchable. Hence, each adjacent pair of pieces in the frame boundary *must* be matched together in any puzzle solution. So, up to rigid transformations, there is only one configuration for these boundary pieces: each piece must be in the same place as it was during the construction. And because the frame spans a width of k and a height of k , there is only one possible position for its boundary.

Consider the set of edges on the boundary of the frame that were colored red or green. There are multiple pieces with red or green sides; therefore, those pieces must be matched with other red or green sides. The only pieces that have red or green edges and do not belong to the frame boundary are the puzzle pieces used to construct each a_i . Therefore, at least one such piece must be placed inside the box we cut out. Consider the set of puzzle pieces used to construct the a_i s. Each piece has four sides: two green sides, and two sides that are colored either red or c_i . None of these colors is unique; therefore, none of the a_i pieces can be adjacent to an empty square. Hence, by induction, the entire $t \times n$ box must be filled with pieces representing the values a_i . An argument similar to the original edge-matching proof shows that we must therefore have a valid 3-Partition. \square