W4231: Analysis of Algorithms

11/30/99

- NP-completeness of Subset Sum, Partition, Minimum Bin Packing.

Subset Sum

The *Subset Sum* problem is defined as follows:

- Given a sequence of integers $a_1, \ldots, a_n$ and a parameter k,
- Decide whether there is a subset of the integers whose sum is exactly k. Formally, decide whether there is a subset $I \subseteq \{1, \ldots, n\}$ such that $\sum_{i \in I} a_i = k$.

We prove that this problem is NP-hard by reduction from Vertex Cover.

Note: Subset Sum is a true *decision problem*, not an optimization problem forced to become a decision problem.

What we have to do

- Start from a graph G and a parameter k.
- Create a sequence of integers and a parameter k'.
- Prove that the graph has vertex cover with k vertices iff there is a subset of the integers that sum to k'.

The Reduction — Overview

We start from a graph $G = (V, E)$ with n vertices. We assume $V = \{1, \ldots, n\}$.

We define integers $a_1, \ldots, a_n$, one for every vertex; and also integers $b_{(i,j)}$, one for every edge $(i, j) \in E$.

We'll define k' later.

Intuitively, we want our instance of subset sum to be such that

- if we have a subset of the a_i and the $b_{(i,j)}$ that sums to k',
- then
  - the subset of the a_i corresponds to a vertex cover C in the graph,
  - and the subset of the $b_{(i,j)}$ corresponds to the edges in the graph such that exactly one of their endpoints is in C.
- Furthermore the construction will force C to be of size k.

Construction

We represent the integers in a matrix. Each integer is a row. the row should be seen as the base-4 representation of the integer.

The first column of the matrix is a special one. It contains 1 for the a_i and 0 for the $b_{(i,j)}$.

Then there is a column for every edge. The column (i, j) has a 1 in a_i, a_j and $b_{(i,j)}$, and all 0s elsewhere.

The parameter k' is defined as

$$k' := k \cdot 4^{|E|} + \sum_{j=0}^{|E|-1} 2 \cdot 4^i$$

Correctness — From Covers to Subsets

Suppose there is a vertex cover C of size k in G.

Then we choose all the integers a_i such that $i \in C$ and all the integers $b_{(i,j)}$ such that exactly one of i and j is in C.

Then, when we sum these integers, doing the operation in base 4, we have a 2 in all digits. In the most significant digit, we are summing 1 $|C| = k$ times.

Correctness — From Subsets to Covers

Suppose we find a subset $C \subseteq V$ and $E' \subseteq E$ such that

$$\sum_{i \in C} a_i + \sum_{(i,j) \in E'} b_{(i,j)} = k'$$

First note that we note that we never have a carry in the $|E|$ less significant digits: operations are in base 4 and there are at most 3 ones in every column.

Since the $b_{(i,j)}$ can contribute at most one 1 in every column, and k' has a 2 in all the $|E|$ less significant digits, it means that for every edge (i,j) C must contain either i or j. So C is a cover.

Every a_i is at least $4^{|E|}$, and k' gives a quotient of k when divided by $4^{|E|}$. So C cannot contain more than k elements.

Partition

- Given a sequence of integers $a_1, \ldots, a_n$.
- Determine whether there is a partition of the number into to subsets such the sum of the elements in one subset is equal to the sum of the elements in the other.

  Formally, determine whether there exists $I \subseteq \{1, \ldots, n\}$ such that $\sum_{i \in I} a_i = (\sum_{i=1}^n a_i)/2$.

Partition is a special case of Subset Sum.

We will prove that it is NP-hard by reduction from Subset Sum.

Reduction

Given an instance of Subset Sum we have to construct an instance of Partition.

Let the instance of Subset Sum have items of size $a_1, \ldots, a_n$ and a parameter k.

Let $A = \sum_{i=1}^n a_i$.

Consider the instance of Partition $a_1, \ldots, a_n, b, c$ where $b = 2A - k$ and $c = A + k$.

Then the total size of the items of the Partition instance is $4A$ and we are looking for the existence of a subset of $a_1, \ldots, a_n, b, c$ that sums to $2A$.

The partition exists iff there exists $I \subseteq \{1, \ldots, n\}$ such that $\sum_i a_i = k$.

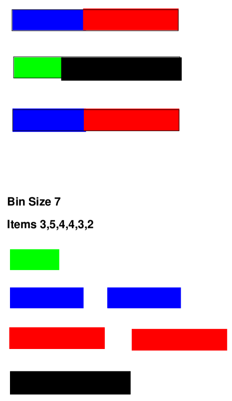
Bin Packing

- Given items of size $a_1, \ldots, a_n$, and given unlimited supply of bins of size B, we want to pack the items into the bins so as to use the minimum possible number of bins.
- Examples of bins/items: tapes and songs; breaks and commercials; bandwith and packets.

- Decision version:
  - Given items of size $a_1, \ldots, a_n$, given bin size B, and parameter k,
  - Determine whether it is possible to pack all the items in k bins of size B.

The problem is NP-hard. Reduction from Partition.

Example



Reduction

Given items of size $a_1, \ldots, a_n$, make an instance of Bin Packing with items of the same size and bins of size $(\sum_i a_i)/2$.

There is a solution for Bin Packing that uses 2 bins if and only if there is a solution for the Partition problem.