A note on the search for maximal pairwise coprime subsets

Prithvijit Chakrabarty

Abstract:
This proves the NP completeness of the search for the largest pairwise coprime subset in a set of integers.

Introduction:

No polynomial time algorithm to factorize integers is known. There is however, an algorithm to factorize an integer into a set of pairwise coprime numbers. This algorithm, called factor refinement, was introduced by Bach, Driscoll and Shallit in [1], and allowed factoring an integer into coprime numbers in quadratic time. The approach was further investigated by Bernstein in [2], where he introduced an algorithm to do same in approximately linear time. Indeed, in many applications, factor refinement can be used instead of integer factorization.

Factor refinement is thus closely linked to integer factorization. This could be useful in investigating the complexity class of integer factorization. Instead of directly looking for a polynomial time algorithm, or proof of NP completeness of factoring, an auxiliary problem, the PC problem is introduced. This is linked to both factor refinement and integer factorization. This paper proves the intermediate result that the PC problem is NP complete. In following work, I will describe the relationship between the PC problem and integer factorization.

Notations:

G is a graph with vertex set V and edge set E

$\bar{G}$ is the complement of graph G

E(G) is the edge set of graph G

e(v1, v2) represents an edge from vertex v_1 to v_2

L :$(V \cup E) \longrightarrow$ Z is the label function that maps vertices and edges to integral labels

The PC problem:

Pairwise coprime set:
A set S is called pairwise coprime if and only if $\gcd(p_1, p_2) = 1 \forall p_1, p_2 \in S$

The PC problem:
Given a set of integers, the PC problem is to find the largest subset that is pairwise coprime.

The factor refinement algorithm generates a pairwise coprime set of numbers, all of which are factors of a given integer (called the coprime base set of the integer in [2]). The PC problem may be related to the integer factorization problem through the fact that every integer can be decomposed into a corresponding coprime base set.

NP Completeness:

The PC problem is proven to be NP complete by reduction from the max clique problem.

Algorithm 2: Reduction from clique to PC
Input: Graph G = (V, E).
Output: Set of vertex labels for the vertices of G

The algorithm maintains a list of labels L for each node and edge. L = {(v_1, p1), (v_2, p2), (v_3, p3)... (e_N, p_N)}, where p_1, p_2, p_3, ..., p_N are distinct primes.

```
**for** e(v1,v2) in E($\bar{G}$) **begin**:
        $L(v_1) \longleftarrow L(v_1) * L(e)$
        $L(v_2) \longleftarrow L(v_2) * L(e)$
end
return L
```

The set of vertex labels L is the input to the corresponding PC problem.

Theorem:
The graph G contains a clique of size 'k' if and only if its corresponding label set contains a pairwise coprime subset with 'k' elements.
Proof:
Consider a subgraph $S_G = \{v_1, v_2, v_3, ..., v_k\}$. The corresponding set of labels will be $S_k = \{L(v_1), L(v_2), L(v_3), ..., L(v_k)\}$. Algorithm 2 ensures that labels $L(v_1)$ and $L(v_2)$ have a common factor if and only if $e(v_1, v_2)$ is in E.

1. S_k is pairwise coprime: This implies that $e(v_i, v_j)$ is in E for all (v_i, v_j) in S_G. Thus, S_G must be a clique.
2. S_k is not pairwise coprime: This implies that there exists at least one pair $L(v_i)$, $L(v_j)$ that are not pairwise coprime and the corresponding edge $e(v_1,v_2)$ is not in E. Thus, S_G will not be a clique if S_k is not pairwise coprime.

Thus, G contains a clique of size 'k' if and only if S contains a pairwise coprime subset with 'k' elements.

Theorem:
The length of the labels generated by algorithm 1 is O(|V|+|E|).
Proof:
The labels of the vertices in G are initialized to |V| distinct primes. To restrict the size of the numbers generated as labels, we use the first |V| prime numbers to label the vertices and the next |E| primes to label the edges before running algorithm 1. Due to the labelling algorithm, the final label of a vertex v_j will have the maximum value if it is connected to all the other vertices in G. In this case,

$$L(v_j) = \prod_{i=|V|+1}^{|V|+|E|} p_i < \prod_{i=0}^{|V|+|E|} p_i.$$

Thus,

$$L(v\ ) < prim(|V|+|E|)\ \forall\ |V|,|E| > 0 \qquad (1)$$

where prim() is the primorial function. The value of the labels is thus bounded by prim(|V|+|E|
Consider N(x) such that N(x) = number of digits in x. If the vertex v_j has the highest label value, the number of digits in $L(v_j)$ will be:

$$N(L(v_j)) = (\lfloor \log(L(v_j)) \rfloor + 1) < \log(prim(|V| + |E|)).$$

The logarithm of the primorial function is the first Chebyshev function $\theta(|V| + |E|)$, which grows linearly, as proved in [3]. Thus, the number of digits in the labels of the vertices of G is generated by the labelling algorithm is O(|V|+|E|).

Theorem:
The reduction of the clique to the PC problem runs in polynomial time.
Proof:
Algorithm 2 traverses the list of edges in E and performs 2 multiplications for every edge. The number of digits of the labels is O(|V|+|E|). Using a simple multiplication algorithm running in quadratic time yields the product in $O((|V|+|E|)^2)$. Thus, the complexity of generating the label set L is $O(|E|.(|V|+|E|)^2)$.

Conclusion and further work:
This proved the NP completeness of the PC problem, which may serve as an auxiliary to classifying the complexity of integer factorization. In future papers, I will investigate the exact

relationship of this problem with integer factorization.

References:

[1] Eric Bach, James Driscoll, Jeffrey Shallit, *Factor Refinement*, in

SODA '90 Proceedings of the first annual ACM-SIAM symposium on Discrete algorithms
Pages 201-211.

[2] Daniel J. Bernstein, *Factoring into coprimes in essentially linear time*, ACM Journal of Algorithms, Volume 54 Issue 1, January 2005.

[3] Hardy, G. H. and Wright, E. M. *The Functions $\theta(x)$ and $\psi(x)$* and *Proof that $\theta(x)$ and $\psi(x)$ are of Order x*, in *An Introduction to the Theory of Numbers*, 5th edition, Pages 340-342, 1979.