Lecture 11 (07.06.2013) Author: Mateusz Baranowski

Randomized compression.

1 Subset Sum

In this section we will use the following notation $[n] = \{1, \ldots, n\}$.

SUBSET SUM
Input: Positive integers $s_1, s_2, \ldots, s_n$ and T.
Question: Does there exist $I \subseteq [n]$ such that $\sum_{i \in I} s_i = T$?

We will consider the problem SUBSET SUM parametrized by n. This problem is clearly in FPT, since we can iterate through all 2^n possible subsets.

Let $p \leq 2^{6n}$ be a random prime number. Consider the following problem.

MODULO SUBSET SUM
Input: Positive integers $s_1, s_2, \ldots, s_n$ and T.
Question: Does there exist $I \subseteq [n]$ such that $\sum_{i \in I} s_i = T \pmod{p}$?

This problem is in NP.

Given an instance of SUBSET SUM $(s_1, \ldots, s_n, T)$ consider the compression to MODULO SUBSET SUM instance $(s'_1, \ldots, s'_n, T') := (s_1 \bmod p, \ldots, s_n \bmod p, T \bmod p)$. This compression has size $\mathcal{O}(n^2)$.

Observation 1. If the SUBSET SUM instance $(s_1, \ldots, s_n, T)$ has a solution then the MODULO SUBSET SUM instance $(s'_1, \ldots, s'_n, T')$ also has a solution.

The prime numbers for which the reverse is not true we will call *bad*. This is only possible when there exists a set $X \subseteq [n]$ such that

$$\sum_{i \in X} s'_i - T' \equiv 0 \pmod{p}.$$

Notice that for a given set X this can only be possible for $\log(n \cdot 2^m)$ prime numbers, where m is the largest number of bits required to represent a number in the instance.

If $2^n \leq m$ then we can use the brutal solution, so assume that $2^n > m$. We have $\log(n2^m) = m + \log n \leq 2^n + \log n \leq 2^{2n}$. Thus the number of bad prime numbers is at most 2^{4n} For large n there is more than 2^{5n} prime numbers smaller than 2^{6n}, so the probability of error is $\leq 2^{-n}$.

Since SUBSET SUM is NP-complete and MODULO SUBSET SUM is in NP, there exists a polynomial reduction from MODULO SUBSET SUM to SUBSET SUM which, applied to our compression, gives a polynomial kernel of the original problem.

Theorem 2 ([2])**.** SUBSET SUM *has an* $\mathcal{O}(n^2)$ *randomized compression and an* $\mathcal{O}(poly(n))$ *randomized kernel.*

2 K-Set Cycle

2.1 Overview

K-Set Cycle
Input: Undirected graph $G = (V, E)$, $K \subseteq V$.
Question: Does there exist a cycle C such that $K \subseteq V(C)$?

This problem is NP-complete since for $K = V$ it is the Hamiltonian Cycle problem. We will consider the parametrization by $|K| = k$. This problem is in FPT since it can be reduced to $k!$ instances of k-Vertex Disjoint Paths. An improved $\mathcal{O}^\star(2^k)$ FPT algorithm was obtained by Björklund et al [1]. During this lecture, we are going to cover the following result of Wahlström.

Theorem 3 ([3])**.** *This problem can be solved in $\mathcal{O}^\star(2^k)$ and there exists a randomized compression.*

We will present a proof of this theorem in the rest of this lecture.

2.2 Tutte Matrix

Definition 4. Let $G = (V, E)$ be an undirected graph with a linear ordering of vertices. We define the *Tutte Matrix* of G as a $|V| \times |V|$ matrix A_T:

$$A_T(u, v) = \begin{cases} x_{u,v} & uv \in E \wedge u < v \\ -x_{v,u} & uv \in E \wedge v > u \\ 0 & \text{otherwise} \end{cases}$$

Definition 5. Let A be an $n \times n$ matrix. Recall that the *determinant* of A is:

$$\det(A) = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) \cdot \prod_{1 \leq i \leq n} a_{i,\pi(i)}$$

Notice that the nonzero polynomials of $\det(A_T)$ correspond to the cycle covers of G.

Lemma 6. $\det(A_T)$ *is the sum of monomials corresponding to the cycle covers consisting exclusively of even cycles. The coefficients of these monomials are* $\pm 2^{\#\text{cycles of length}>2}$.

Proof. We can pair up cycle covers by reversing the direction of the odd length cycle with the smallest vertex. The reversing of an odd cycle changes only the sign of the monomial, thus monomials representing covers containing odd cycles will cancel each other out.

Reversing an even cycle doesn't change the monomial and there is $2^{\#\textit{cycles of length}>2}$ orientations of a cycle cover. □

Lemma 7. *Given an instance $I = (G, K)$ of k-Set Cycle we can find an equivalent instance $I' = (G', K')$ such that $|K'| = |K|$, K' is an an independent set in G' and $\forall_{r \in K'} \deg_{G'}(v) = 2$.*

Proof. First we subdivide the edges of G so that there are no edges between vertices in K. Next for each vertex $v_i \in K$ we create two vertices x_i, y_i which we connect with $\mathcal{N}[v_i]$. Finally, we make x_i and y_i the sole neighbours of v_i. □

From this point we assume that the elements of matrices belong to a field of size 2^r and characteristic 2 (this means that for each element x we have $x + x = 0$).

$$B_T(u,v) := \begin{cases} A_T(u,v) & u \neq v \vee u \in K' \\ 1 + A_T(u,v) & \text{otherwise} \end{cases}$$

This transformation allows the use of cycles of length 1 for vertices not in K' in the cycle covers. The determinant of B_T will contain monomials representing cycle covers consisting of cycles of length greater than one containing vertices of K' and a number of cycles (possibly of length 1) for vertices from $V \setminus K'$ (if such a cover exists).

Let $K' = \{v_1, \ldots, v_k\}$.

$$C_T(u,v) = \begin{cases} 0 & u = x_1 \wedge v = v_1 \\ B_T(u,v) & \text{otherwise} \end{cases}$$

This allows us to select an orientation of the cycle containing v_1. Since now only one will be valid, this monomial won't cancel out in our future calculations.

2.3 FPT algorithm

For a function $f : \{v_2, v_3, \ldots, v_k\} \to \{0,1\}$ we define a matrix D_f as follows

$D_f := C_T$
for all $2 \leq i \leq k$ **do**
  if $f(i) = 0$ **then**
    $D_f(x_i, v_i) := 0$
    $D_f(v_i, y_i) := 0$
  else
    $D_f(y_i, v_i) := 0$
    $D_f(v_i, x_i) := 0$
  end if
end for

Note that f selects an orientation for cycles containing vertices of K'.

Define $P := \sum_{f:\{v_2,\ldots,v_l\}\to\{0,1\}} \det(D_f)$. The following lemma shows that P is symbolically nonzero, iff our instance admits a solution.

Lemma 8. *$P \not\equiv 0$ iff (G', K') has a solution.*

Proof. "⇐": the monomial representing the cycle cover consisting of exactly one cycle of length > 1, the cycle containing the vertex set K', and cycles of length 1. Since we forced the orientation of this cycle, it will appear only once in $\sum_f \det(D_f)$. Thus it will not be canceled out.

"⇒": We will show that every cycle cover not containing a cycle containing all vertices of K' will be added an even number of times. Consider the cycle, which

a) does not contain v_1,

b) contains an element from $K' \setminus \{v_1\}$,

c) break ties taking the cycle containing the smallest index vertex from K'.

Due to the construction of our matrix this cycle has length ≥ 3. We can pair up this cycle cover with a cover obtained by reversing this cycle. Thus, due to the choice of a field with characteristic 2, it will cancel out in P. □

Obviously we are not able to compute the symbolic determinant, as it may be of exponential size. For this reason, we use the Zippel-Schwartz lemma which appeared in previous lectures.

Lemma 9 (Zippel-Schwartz)**.** *Let $p(x_1, \ldots, x_n)$ be a nonzero polynomial of degree d over a field $\mathbb{F}$. Let us draw z_i uniformly at random from a subset $X \subseteq \mathbb{F}$ of size N. Then $p(z_1, \ldots, z_n) = 0$ with probability at most $\frac{d}{N}$.*

Therefore in order to check whether P is nonzero it is enough to evaulate it in a single point, which leads to a computation of 2^k determinants, thus the time complexity is $\mathcal{O}^\star(2^k)$.

2.4 Compression

Let $f, f' : \{v_2, v_3, \ldots, v_k\} \to \{0, 1\}$. Notice that the matrices D_f and $D_{f'}$ have very little differences. Let $a_2, \ldots, a_k$ be new variables. We define the matrix E as follows:

$E := C_T$
for all $2 \leq i \leq k$ **do**
  $E(x_i, v_i) := E(x_i, v_i) \cdot a_i$
  $E(v_i, y_i) := E(v_i, y_i) \cdot a_i$
  $E(y_i, v_i) := E(y_i, v_i) \cdot (1 - a_i)$
  $E(v_i, x_i) := E(v_i, x_i) \cdot (1 - a_i)$
end for

Let $g : \{a_2, \ldots, a_k\} \to \{0, 1\}$. We will note as $g[E]$ the matrix E with values assigned to variables a_i according to g. Notice that if $\forall_i f(v_i) = g(a_i)$ then $g[E] = D_f$.

Observation 10. (G', K') has a solution iff

$$\sum_{g:\{a_2,\ldots,a_k\}\to\{0,1\}} \det(g[E]) = P \not\equiv 0$$

Thus it suffices to compress E. Notice that variables a_i appear only in $\mathcal{N}[K']$, thus we can rearrange the matrix so that we get:

$$\begin{pmatrix} E_1 & R_1 \\ R_2 & E_2 \end{pmatrix}$$

(1) Every element of E_1 is of the form $ba_i + c$ for $b, c \in \mathcal{F}$.

(2) No element outside of E_1 contains variables a_i.

Observation 11. With high probability the determinant of E_2 is nonzero.

Using this observation we can apply the Gaussian elimination on E_2 to get a diagonal matrix which we use to transform our matrix to:

$$F = \begin{pmatrix} F_1 & 0 \\ 0 & F_2 \end{pmatrix}$$

Now, since F_2 doesn't contain variables a_i, we can remember only F_1 and $\det(F_2)$. F_1 is a matrix of size $3k \times 3k$ and each element is of the form $ba_i + c$, where $b, c \in \mathcal{F}$. Thus the number of elements to remember is polynomial.

The only question left is how to select the size of the field so that we get a polynomial compression with high probability. We split this problem into two cases:

- $2^k \leq |V| \Rightarrow$ we use the FPT algorithm, which works in poly($|V|$).
- $2^k > |V| \Rightarrow$ we take the size of the field $2^{c \log |V|}$. From the Zippel-Schwartz lemma we get that the probability of failure is $\leq \frac{1}{|V|^{c-1}}$.

Open problem 1. How to construct a polynomial kernel for k-Set Cycle?

Open problem 2. Is it possible that there exists a polynomial compression but not a polynomial kernelization?

References

[1] Andreas Björklund, Thore Husfeldt, and Nina Taslaman. Shortest cycle through specified elements. In *SODA*, pages 1747–1753, 2012.

[2] Danny Harnik and Moni Naor. On the compressibility of *np* instances and cryptographic applications. *SIAM J. Comput.*, 39(5):1667–1713, 2010.

[3] Magnus Wahlström. Abusing the tutte matrix: An algebraic instance compression for the k-set-cycle problem. In *STACS*, pages 341–352, 2013.