26.3 Sum-free sets: the algorithmic aspect

In previous sections we considered two general approaches toward derandomizing of probabilistic proofs. In this section we will give one example to demonstrate that sometimes the desired polynomial-time algorithm is hidden in the existence proof *itself.*

A subset B of an additive group is called *sum-free* if $x + y \notin B$ for all $x, y \in B$. Erdős (1965) and Alon and Kleitman (1990) have proved that every set $A = \{a_1, \ldots, a_N\}$ of integers has a sum-free subset B, with $|B| > N/3$. The proof is probabilistic (see Theorem 20.2) and the question was whether there exists a deterministic algorithm for the selection of such a subset B, which runs in time polynomial in the (binary) size of the problem, that is in $\ell \rightleftharpoons \sum_{i=1}^{N} \log_2 |a_i|$.

Kolountzakis (1994) has shown that, with a slight modification, the proof of Theorem 20.2 can be transformed to such an algorithm.

For a prime p let (as before) $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$ be the field of the integers mod p, and let $\mathbb{Z}_p^{\times} = \{1, \ldots, p-1\}$ be the corresponding multiplicative group in $\mathbb{Z}_p$.

Theorem 26.6 (Kolountzakis 1994)**.** *Let $p = 3k+2$ be a prime number and $w(x)$ a nonnegative function defined on $\mathbb{Z}_p^{\times}$. Define $W \rightleftharpoons \sum_{x \in \mathbb{Z}_p^{\times}} w(x)$ and assume $W > 0$. Then there is a sum-free subset E of $\mathbb{Z}_p^{\times}$ for which*

$$\sum_{x \in E} w(x) > \frac{1}{3} W. \tag{26.3}$$

Proof. Write $S = \{k+1, k+2, \ldots, 2k+1\}$, and observe that S is a sum-free subset in $\mathbb{Z}_p$ and $|S| > (p-1)/3$. Let the random variable $\mathbf{t}$ be uniformly distributed in $\mathbb{Z}_p^{\times}$, and write $f(\mathbf{t}) \rightleftharpoons \sum w(x)$, where the sum is over all x for which $x \cdot \mathbf{t} \in S$, and the product $x \cdot \mathbf{t}$ is computed in $\mathbb{Z}_p$. Since $\mathbb{Z}_p^{\times}$ is a multiplicative group, we have

$$\mathrm{E}\left[f(\mathbf{t})\right] = W \cdot (|S|/(p-1) > W/3.$$

By the pigeonhole property of the expectation, there is some $t \in \mathbb{Z}_p^{\times}$ for which $f(t) > W/3$. Define $E \rightleftharpoons t^{-1}S$. This set is sum-free and (26.3) true for it. □

We now turn this proof into an algorithm. Given a set $A = \{a_1, \ldots, a_N\}$ of integers of (binary) size $\ell \rightleftharpoons \sum_{i=1}^{N} \log_2 |a_i|$, our goal is to find a sum-free subset B, with $|B| > N/3$, in time polynomial in ℓ. We assume that ℓ is large.

First, observe that the number of prime factors of an integer x is at most $\log_2 x$. This means that the number of prime factors which appear in the factorization of *any* element of A is at most ℓ. The Prime Number Theorem says that for every pair b, c of relatively prime positive integers, the number of primes $p \leqslant x$ such that p is of the form $p = bk + c$, asymptoticaly equals to $x/(\varphi(b) \cdot \ln x)$, where $\varphi(b) = |\{y \in \mathbb{Z}_b : \gcd(y, b) = 1\}|$ is the Euler totient function. In our case $b = 3$ and $c = 2$; hence, $\varphi(b) = 2$. Thus, there is a prime p of the form $p = 3k + 2$, not greater than $3\ell \log_2 \ell$, which does not divide any member of A.

Define now

$$w(x) \rightleftharpoons |\{t \in A : t = x \bmod p\}| .$$

Since p does not divide any member of A, we have $W = N$ and, using Theorem 26.6, we can find a sum-free subset $E \subseteq \mathbb{Z}_p^{\times}$ for which the set

$$B = \{t \in A : t \bmod p \in E\}$$

has more than $N/3$ elements. This set B is sum-free since $x + y = z$ for some $x, y, z \in B$ would imply $x + y = z \bmod p$ and E would not be sum-free.

In summary, the steps of our algorithm are the following.

1. Compute all primes up to $3\ell \log_2 \ell$.
2. Find a prime $p = 3k + 2$ which divides no element of A.
3. Compute the values $w(x)$ for all $x \in \mathbb{Z}_p^{\times}$.
4. Find by exhaustive search a $t \in \mathbb{Z}_p^{\times}$ for which $f(t) > N/3$ (Theorem 26.6 guarantees that such t exists) and compute the set $E = t^{-1}S$.
5. Construct the set $B = \{t \in A : t \bmod p \in E\}$.

It is easy to verify (do this!) that all these steps can be carried out in time polynomial in ℓ.

Exercises

26.1.$^{(!)}$ Use the method of conditional probabilities to derandomize the proof of Theorem 18.1 and Theorem 18.2.

26.2.$^{-}$ Let $G = (V, E)$ be a graph with $n = 2m$ vertices. Improve the lower bound $|E|/2$ on the size of a cut in G (proved in Theorem 26.1) to $|E| \geqslant m/(2m - 1)$.

Hint: Follow the argument of Theorem 26.1 with another probability space: choose $\mathbf{U} \subseteq V$ uniformly from among all m-element subsets of V. Observe that then any edge has probability $m/(2m - 1)$ of being crossing.

26.3.$^{-}$ Let $\mathbf{r}$ be a random vector uniformly distributed in $\mathbb{F}_2^d$. With each vector $a \in \mathbb{F}_2^d$ associate a random variable $X_a = \langle a, \mathbf{r} \rangle$ whose value is the scalar product over $\mathbb{F}_2$ of this vector with $\mathbf{r}$. Show that these random variables are 2-wise independent. *Hint*: Exercise 17.2.

26.4. Let $\log m = o(\sqrt{n})$, $m > 4$, and let H be an $m \times n$ 0-1 matrix, the average density of (i.e., the average number of 1's in) each row of which does not exceed p, $0 \leqslant p < 1$. Show that then, for every constant $\delta > 0$, there is an $m \times t$ submatrix H' of H such that $t = O(\log m/\delta^2)$ and each row of H' has average density at most $p + \delta$.