Fast evaluation of the prime generating function in $\mathbb{F}_2[x]$

by Polymath4

September 22, 2009

1 Notation

Given an interval $[a, b]$, by $|[a, b]|$ we will mean its length, which is $b - a$.

2 Introduction

Let

$$c \ := \ 1/100,$$

and let $\delta > 0$ to be any constant such that for any quadratic polynomial

$$\ell(j) \ = \ a_0 j^2 + a_1 j + a_2, \text{ where } |a_0|, |a_1|, |a_2| \ < \ n^{O(1)},$$

we can evaluate

$$\sum_{j \leq J} x^{\ell(j)} \pmod{2, g(x)}$$

in time at most

$$J^{1-\delta} (\log n)^{O(1)} (\deg(g))^{O(1)}.$$

Finally, let $\varepsilon > 0$ be some constant that will be chosen later in terms of c and δ.

Given a positive integer n, using Odlyzko's algorithm we know that we can locate, in time

$$n^{1/2 - \varepsilon + o(1)},$$

an interval

$$I \;:=\; [z, z + z^{1/2+\varepsilon}] \;\subseteq\; [n, 2n],$$

containing at least one prime. Our job will be to show that for $\varepsilon > 0$ small enough (but independent of n) and $n > n_0(\varepsilon)$, we can compute the generating function

$$h(x) \;:=\; \sum_{m \in I} \tau^*(m) x^m \pmod{2, g(x)},\; \tau^*(m) \;=\; |\{d|m \;:\; d \leq \sqrt{n}\}|$$

in time at most about

$$n^{1/2-\delta c/4+O(\varepsilon)}(\deg(g(x)))^{O(1)}. \tag{1}$$

Using this in combination with the ideas in my note on how to leverage a power savings in the computation of divisor sums to obtain a power savings in the computation of the parity of $\pi(x)$, we will show how to compute

$$\sum_{\substack{p \in I \\ p \text{ prime}}} x^p \pmod{2, g(x)},$$

in time $z^{1/2-\Omega(c)}$. The details of how to do this can be found in section 4.

3 The algorithm

3.1 The basic approach

Let

$$f_d(x) \;:=\; \sum_{\substack{m \in I \\ d|m}} x^m.$$

Then, we have that

$$h(x) \;=\; \sum_{d \leq \sqrt{n}} f_d(x) \;=\; \sum_{d \leq n^{1/2-c/4}} f_d(x) + \sum_{n^{1/2-c/4} < d \leq n^{1/2}} f_d(x).$$

Let $r(x)$ denote this last sum over $d \in [n^{1/2-c/4}, n^{1/2}]$. It is this sum which we will show can be evaluated quickly; to evaluate the sum over $d \leq n^{1/2-c/4}$ we just use the geometric series formula (or the identity $(1+X)^{2^j-1} = 1 + X + \cdots + X^{2^j-1} \in \mathbb{F}_2[x]$).

First, we partition $r(x)$ as follows

$$r(x) \;=\; \sum_{k=1}^{K} \sum_{d \in D_k} f_d(x),$$

where

$$K \;:=\; \lfloor n^{1/2-c} \rfloor,$$

and where we let the intervals $D_1, ..., D_K$ be consecutive (disjoint) intervals of width $(n^{1/2} - n^{1/2-c/4})/K$ that cover $[n^{1/2-c/4}, n^{1/2}]$.

Our goal now is to show that each of the sums

$$S_k(x) \;:=\; \sum_{d \in D_k} f_d(x)$$

can be computed in time at most $|D_k| n^{-\delta c/4 + O(\varepsilon)}$, which clearly would lead to our sought after running time given in (1).

3.2 The fraction u/v

To compute $S_k(x) \pmod{2, g(x)}$ quickly, we begin by letting γ_1, γ_2 satisfy

$$D_k \cap \mathbb{Z} \;=\; [\gamma_1,\ \gamma_2] \cap \mathbb{Z},\ \text{where } \gamma_1, \gamma_2 \in \mathbb{Z},$$

and then we let θ_1 denote the smallest integer such that

$$\gamma_2 \theta_1 \;\in\; I,$$

and we let θ_2 denote the largest integer such that

$$\gamma_2 \theta_2 \;\in\; I.$$

Next, using the continued fraction algorithm we find a rational approximation

$$u/v,\ 1 \leq v \leq n^{c/2}$$

to

$$\gamma_2/\theta_2,$$

so that

$$\left| \frac{\gamma_2}{\theta_2} \;-\; \frac{u}{v} \right| \;\ll\; \frac{1}{v n^{c/2}}.$$

Let us assume initially that u/v is in fact an *under-approximation* to γ_2/θ_2.

3.3 Partitioning D_k

For various technical reasons appearing at the end of subsection 3.5, D_k is a little too big to work with, so we will need to chop it up into

$$t \ := \ \lfloor n^{c/2}/v \rfloor \ \ \text{equal} - \text{sized pieces.}$$

Call these intervals

$$D_k(1), \ ..., \ D_k(t),$$

with $D_k(t)$ denoting the interval with the largest elements of D_k. Note that

$$|D_k(i)| \ = \ |D_k|/t \ \gg \ n^c/t \ \gg \ vn^{c/2}.$$

3.4 Computing the sum over d quickly

Observe that

$$\begin{aligned} S_k(x) \quad &= \quad \sum_{a=0}^{v-1} \ \sum_{\substack{d \in D_k \\ d \equiv a \pmod v}} f_d(x) \\ &= \quad \sum_{h=1}^{t} \sum_{a=0}^{v-1} \ \sum_{\substack{d \in D_k(h) \\ d \equiv a \pmod v}} f_d(x). \end{aligned}$$

What we will now do is fix $h = 1, ..., t$ and $a = 0, ..., v-1$, and then show that we can evaluate the corresponding inner sum very quickly.

It will turn out that computing each of these inner sums quickly (for different values of a and h) will be virtually identical, so we will only bother to focus on the case

$$a \ \equiv \ \gamma_2 \pmod v, \ \text{and } h \ = \ t. \tag{2}$$

3.5 A generic sum over d

The contribution of the inner sum terms in $S_k(x)$ represented by the case (2) will basically be

$$\sum_{0 \le j \le (\gamma_2 - \gamma_1)/tv} f_{\gamma_2 - jv}(x). \tag{3}$$

Among the terms appearing in this sum will be

$$\sum_{0 \le j \le J} x^{(\gamma_2 - jv)(\theta_2 + ju)}, \tag{4}$$

where J is maximal, subject to the constraints

$$(\gamma_2 - jv)(\theta_2 + ju) \ \in \ I, \text{ and } \gamma_2 - jv \in D_k(t). \tag{5}$$

Of course there will be other sums like (4) that contribute to (3), and those will be discussed below.

First, let us see how large J is which guarantees that

$$(\gamma_2 - jv)(\theta_2 + ju) \ \in \ I, \ 0 \le j \le J. \tag{6}$$

To work this out, note that upon expanding it out we get

$$\gamma_2\theta_2 + j(u\gamma_2 - v\theta_2) + j^2 uv.$$

Now, $\gamma_2\theta_2$ is certainly in I, and is in fact in the top half of I. And, from the fact that u/v is a good under-approximation to γ_2/θ_2, we find that

$$j(u\gamma_2 - v\theta_2) < 0$$

and

$$|j(u\gamma_2 - v\theta_2)| \ = \ |(jv\gamma_2)(u/v - \theta_2/\gamma_2)| \ \ll \ j\gamma_2 n^{-c/2} \ < \ jn^{1/2-c/2}. \tag{7}$$

So, we have that

$$J \ \ge \ n^{c/2}/2;$$

and, the second constraint of (5) amounts to having

$$J \ \le \ |D_k(t)|/v \ \ll \ n^{c/2};$$

so, up to a constant factor, J is of size $n^{c/2}$.

Note that since the exponent is quadratic in j, we can apply the algorithm from previous postings to evaluate it in time at most, say,

$$J^{1-\delta+o(1)}(\deg(g(x)))^{O(1)};$$

and this savings of $J^{-\delta}$ amounts to a factor $n^{-\delta c/2}$ overall savings to the running time.

The idea for the other $d \equiv \gamma_2 \pmod v$ will be the same, as it will for those $d \equiv a \pmod v$, $a \neq \gamma_2$, and what will make all this possible is the fact that the corresponding ratios γ_2'/θ_2' (in place of γ_2/θ_2 as above) that arise, come very close to γ_2/θ_2, and therefore very close to u/v. There is a small issue, though, with u/v flipping from being an under-approximation to γ_2/θ_2 to being an over-approximation to γ_2'/θ_2', and this will be discussed in subsection 3.7 below.

3.6 The contribution of the other $d \in D_k(t)$ satisfying $d \equiv \gamma_2 \pmod v$

Not only do we have that

$$(\gamma_2 - jv)(\theta_2 + jv) \ \in \ I,$$

for various different j, but we also have that for a fixed i and j in certain ranges,

$$(\gamma_2 - jv)(\theta_2 + jv - i) \ \in \ I$$

Note that such products still leave the smaller factor $d = \gamma_2 - jv$ satisfying $d \equiv \gamma_2 \pmod v$.

There will actually be two kinds of i that we consider here: the normal ones where $i \geq 0$, and the exceptional one where $i < 0$.

3.6.1 The contribution of the normal values of i

And how big can i be here? Well, since I has width

$$z^{1/2+\varepsilon} \ \ll \ n^{1/2+\varepsilon},$$

and since the $\gamma_2 - jv \geq z^{1/2-c/4}$, the range for i has size at most

$$n^{c/4} z^{\varepsilon} \ \ll \ n^{c/4+\varepsilon}.$$

All these values of i, except for the largest in the range, which we will denote by i_0, will give us that

$$\gamma_2(\theta_2 - i) \ \in \ [z + z^{1/2-\varepsilon}, \ z + z^{1/2+\varepsilon}]. \tag{8}$$

Although this range is not the "top half" of I, that we used before in subsection 3.5 to give a lower bound on the size of J, it is still good enough (because we get to choose $\varepsilon > 0$ as small as needed relative to c). Let us calculate the size of J for each of these i; that is, let us calculate a lower bound for the largest value of j which guarantees that

$$(\gamma_2 - jv)(\theta_2 + ju - i) \ \in\ I, \text{ and } (\gamma_2 - jv)\ \in\ D_k(t). \tag{9}$$

First, expanding out this product, we find that it is

$$\gamma_2\theta_2 - j(v\theta_2 - u\gamma_2) - j^2uv - i(\gamma_2 - jv).$$

Applying now (8), we find that this lies in I provided that

$$|j(v\theta_2 - u\gamma_2 - iv)| \ \ll\ z^{1/2-\varepsilon}.$$

We can basically ignore the iv term here, so that upon applying (7) we have that

$$J \ \gg\ \frac{z^{1/2-\varepsilon}}{|v\theta_2 - u\gamma_2|} \ \gg\ \frac{n^{1/2-\varepsilon}}{n^{1/2-c/2}} \ =\ n^{c/2-\varepsilon}.$$

And, the second constraint in (9) gives us

$$J \ \ll\ n^{c/2}.$$

Since we get to choose $\varepsilon > 0$ as small as desired, the savings we will get by working with these J terms together, instead of applying the geometric series formula to compute $S_k(x)$, will be significant. Indeed, for each i under consideration, with the exception of $i = i_0$ (to be handled below), we can evaluate

$$\sum_{j \le J} x^{(\gamma_2 - jv)(\theta_2 + ju - i)} \pmod{2, g(x)}$$

in time at most

$$J^{1-\delta+o(1)}(\deg(g(x)))^{O(1)}. \tag{10}$$

Next, we handle the case where $i = i_0$: what we do for these terms is we consider the associated value of J. If $J < n^{c/4}$, then of course we can evaluate

$$\sum_{j \le J} x^{(\gamma_2 - jv)(\theta_2 + ju - i_0)} \pmod{2, g(x)}$$

in time at most

$$n^{c/4+o(1)}(\deg(g(x)))^{O(1)}.$$

And on the other hand if $J > n^{c/4}$, then we apply our Strassen idea (or FFTs, if they are available), to compute the sum using fewer than (10) computations.

3.6.2 The exceptional i

Finally, there is one more set of terms left to consider, and these are caused by the fact that some $d \in D_k(t)$ with $d \equiv \gamma_2 \pmod{v}$ may divide more integers in I than does γ_2. The source of the problem here is that these products take the form

$$(\gamma_2 - jv)(\theta_2 + i' + ju), \text{ where } i' > 0,$$

where note that this product strays outside of I when $j = 0$, though may be within I for j somewhat larger. Fortunately for us, the only way this can happen is if

$$0 \leq i' \ll 1.$$

To see this, first note that if i' is sufficiently large, even if we took j as large as $n^{c/2}$, the largest allowed in order to keep

$$(\gamma_2 - jv) \in D_k(t),$$

we would get that our product is at least

$$\gamma_2\theta_2 - j(v\theta_2 - u\gamma_2) - j^2uv + i'(\gamma_2 - jv). \tag{11}$$

And then, applying (7), along with the fact that $\gamma_2\theta_2$ is the largest multiple of γ_2 lying in I, making

$$\gamma_2\theta_2 \geq z + z^{1/2+\varepsilon} - \gamma_2,$$

we deduce that (11) exceeds

$$(z + z^{1/2+\varepsilon} - \gamma_2) - \kappa\gamma_2 - j^2uv + i'(\gamma_2 - jv) > z + z^{1/2+\varepsilon},$$

for some $\kappa > 0$. Clearly, $i' \ll 1$ in order for this to lie in I (and in particular, to be smaller than $z + z^{1/2+\varepsilon}$).

And the contributions of these $O(1)$ values of i' to $S_k(x)$ can be sped up using the Strassen algorithm, just like with the normal values of i in the previous sub-subsection.

3.7 Handling flips from under- to over-approximations (and vice versa)

Basically, we will show that this occurs so rarely that we can just compute the contribution to our generating function of these $d \in D_k$ using the geometric series identity approach, and it will not much affect our overall running time.

3.8 What if u/v is an over-approximation to γ_2/θ_2?

4 Leveraging fast divisor sum computations to quickly compute prime generating functions for $[z, z + z^{1/2+\varepsilon}]$

.