A PSEDUO-TWIN PRIMES THEOREM

ALEX V. KONTOROVICH

1. Introduction

The Twin Prime Conjecture states that there are infinitely many primes p such that $p+2$ is also prime. A refined version of this conjecture is that $\pi_2(x)$, the number of prime twins lying below a level x, satisfies

$$\pi_2(x) \sim C\frac{x}{\log^2 x},$$

as $x \to \infty$, where $C \approx 1.32032\ldots$ an arithmetic constant.

The best result towards the Twin Prime Conjecture is Chen's theorem [Che73], that there are infinitely many primes p for which $p+2$ is either prime or the product of two primes. This statement exhibits the "parity problem," that often methods in sieve theory cannot distinguish between sets having an even or odd number of factors. Vinogradov's resolution [Vin37] of the ternary Goldbach problem introduced the idea that estimating certain bilinear forms can sometimes break this parity, and there have since been many impressive instances of this phenominon, see e.g. [FI98, HB01].

In this note, we aim to illustrate parity breaking in a simple, self-contained example. Consider an analogue of the twin prime conjecture, but instead of intersecting two copies of the primes, we intersect one copy of the primes with a set which analytically mimics the primes. Let $\mathrm{iL}(x) \sim x\log x$ denote the inverse to the logarithmic integral function,

$$\mathrm{Li}(x) := \int_2^x \frac{dt}{\log t}.$$

Definition 1.1. *Let $\tilde{\pi}(x)$ denote the number of primes $p \leq x$ such that $p = \lfloor \mathrm{iL}(n) \rfloor$ for some integer n.*

Here $\lfloor\cdot\rfloor$ is the floor function, returning the largest integer not exceeding its argument. Our main goal is to demonstrate

---

Date: April 23, 2009.

2000 *Mathematics Subject Classification.* 11L07, 11L15, 11L20, 11C08, 11N32, 11N35 , 11N36.

Key words and phrases. Exponential Sums, Bilinear Forms, Piatetski-Shapiro Primes.

Theorem 1. *As* $x \to \infty$,

$$\hat{\pi}(x) \sim \frac{x}{\log^2 x}.$$

Notice that the constant above is 1, that is, there is no arithmetic interference. This theorem follows also from the work of Leitmann [Lei77], generalizing Piatetski-Shapiro's theorem [Pu53]. Our aim is to give a short proof of this statement from scratch.

In §2, we reduce Theorem 1 to an exponential sum over the primes. We devote §3 to breaking the sum into ones of "Type I" and "Type II," estimating these separately. We reserve some technical calculations for the appendix.

Acknowledgements. The author wishes to thank Peter Sarnak for suggesting this problem, and generously lending of his time. Thanks also to Dorian Goldfeld and Patrick Gallagher for enlightening conversations, and to Steven J. Miller and Tim Browning for helpful remarks and corrections to an earlier draft.

2. Reduction to Exponential Sums

We follow standard methods, which we include here for completeness. If $p = \lfloor \mathrm{iL}(n) \rfloor$ then $p \le \mathrm{iL}(n) < p+1$, or equivalently, $\mathrm{Li}(p) \le n < \mathrm{Li}(p+1)$. The existence of an integer in the interval $[\mathrm{Li}(p), \mathrm{Li}(p+1))$ is indicated by the value $\lfloor \mathrm{Li}(p+1) \rfloor - \lfloor \mathrm{Li}(p) \rfloor$, so we have

$$\hat{\pi}(x) = \sum_{p \le x} \lfloor \mathrm{Li}(p+1) \rfloor - \lfloor \mathrm{Li}(p) \rfloor.$$

Write $\lfloor \theta \rfloor = \theta - \psi(\theta) - \frac{1}{2}$, where ψ is the shifted fractional part

$$\psi(\theta) := \{\theta\} - \frac{1}{2} \in [-\frac{1}{2}, \frac{1}{2}).$$

One readily computes the Fourier expansion of ψ; in truncated form, it is

$$\psi(\theta) = \sum_{0<|h|\le H} c_h \, e(\theta h) + O(\frac{1}{H}), \tag{2.1}$$

where $c_h \ll \frac{1}{h}$, and $e(x) = e^{2\pi i x}$. In the above, H is a parameter which we will choose later (we will eventually set $H = \log^2 N$).

So we have:

$$\hat{\pi}(x) = \sum_{p \le x} (\mathrm{Li}(p+1) - \mathrm{Li}(p)) + \sum_{p \le x} (\psi(\mathrm{Li}(p)) - \psi(\mathrm{Li}(p+1))).$$

Since $\mathrm{Li}'(x) = \frac{1}{\log x}$, we use the Taylor expansion:

$$\mathrm{Li}(p+1) = \mathrm{Li}(p) + \frac{1}{\log p} + O(\frac{1}{p\log^2 p})$$

to get:

$$\hat{\pi}(x) = \sum_{p\leq x} \frac{1}{\log p} + \sum_{p\leq x} (\psi(\mathrm{Li}(p)) - \psi(\mathrm{Li}(p+1))) + O(1).$$

By partial summation and a crude form of the prime number theorem,

$$\begin{aligned} \sum_{p\leq x} \frac{1}{\log p} &= \int_2^x \frac{d\pi(t)}{\log t} = \frac{\pi(x)}{\log x} + O\left(\int_2^x \frac{\pi(t)}{t\log^2 t} dt\right) \\ &= \frac{x}{\log^2 x} + O\left(\frac{x}{\log^3 x}\right). \end{aligned}$$

Therefore to prove Theorem 1, it suffices to show that

$$\sum_{p\leq x} \psi(\mathrm{Li}(p)) - \psi(\mathrm{Li}(p+1)) \ll \frac{x}{\log^3 x}. \tag{2.2}$$

Equivalently, split the sum into dyadic segments and apply partial summation to reduce (2.2) to the statement

$$\Sigma := \sum_{N<n\leq N_1\leq 2N} \Lambda(n) \left(\psi(\mathrm{Li}(n)) - \psi(\mathrm{Li}(n+1))\right) \ll \frac{N}{\log^2 N},$$

with $N \ll x$. Here Λ is the von Mangoldt function:

$$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^k \text{ is a prime power} \\ 0 & \text{otherwise.} \end{cases}$$

Using (2.1) we write the sum above as $\Sigma = \Sigma_1 + O(\Sigma_2)$ where

$$\Sigma_1 := \sum_n \Lambda(n) \sum_{0<|h|\leq H} c_h \left(e(h\,\mathrm{Li}(n)) - e(h\,\mathrm{Li}(n+1))\right)$$

and

$$\Sigma_2 := \frac{1}{H} \sum_{n\sim N} \Lambda(n).$$

It is clear by the Prime Number Theorem that $\Sigma_2 \ll N/H$, so choosing $H = \log^2 N$ dispenses with the error.

On writing $\phi_h(x) = 1 - e(h(\mathrm{Li}(x+1) - \mathrm{Li}(x))$ and by partial summation, we see that

$$\begin{aligned}
\Sigma_1 &\ll \sum_{1\le h\le H} h^{-1} \left| \sum_{N<n\le N_1} \Lambda(n)\phi_h(n) e(h\,\mathrm{Li}(n)) \right| \\
&\ll \sum_{1\le h\le H} h^{-1} \left| \phi_h(N_1) \sum_{N<n\le N_1} \Lambda(n) e(h\,\mathrm{Li}(n)) \right| \\
&\quad + \int_N^{N_1} \sum_{1\le h\le H} h^{-1} \left| \frac{\partial \phi_h(x)}{\partial x} \sum_{N<n\le x} \Lambda(n) e(h\,\mathrm{Li}(n)) \right| dx \\
&\ll \frac{1}{\log N} \max_{N_2 \le 2N} \sum_{1\le h\le H} \left| \sum_{N<n\le N_2} \Lambda(n) e(h\,\mathrm{Li}(n)) \right|.
\end{aligned}$$

Here we used the bounds

$$\phi_h(x) \ll h(\mathrm{Li}(x+1) - \mathrm{Li}(x)) \ll \frac{h}{\log N}$$

and

$$\frac{\partial \phi_h(x)}{\partial x} \ll h\left(\frac{1}{\log(x+1)} - \frac{1}{\log(x)} \right) \ll \frac{h}{N \log^2 N}$$

for $N \le x \le 2N$. We have thus reduced Theorem 1 to the estimate:

$$S := \sum_{0<h\le H} \left| \sum_{N<n\le N_2\le 2N} \Lambda(n) e(h\,\mathrm{Li}(n)) \right| \ll \frac{N}{\log N}, \tag{2.3}$$

which we prove in the next section.

3. Bilinear Forms

Our goal in this section is to demonstrate (2.3). We will actually prove more; instead of a log savings, we will save a power:

Theorem 2. *For S defined in* (2.3)*, we have*

$$S \ll N^{21/22+\epsilon}.$$

Fix u and v, parameters to be chosen later, and let $F(s) = \sum_{1\le n\le v} \Lambda(n) n^{-s}$ and $M(s) = \sum_{1\le n\le u} \mu(n) n^{-s}$, where μ is the Möbius function:

$$\mu(n) = \begin{cases} (-1)^k & \text{if } n \text{ is the product of } k \text{ distinct primes} \\ 0 & \text{if } n \text{ is not square-free.} \end{cases}$$

The functions F and M are the truncated Dirichlet polynomials of the functions $-\zeta'/\zeta$ and $1/\zeta$, respectively, where $\zeta(s)$ is the Riemann zeta function. Notice, for instance, that

$$\frac{\zeta'}{\zeta}(s) + F(s) = -\sum_{n>v} \Lambda(n) n^{-s}.$$

Comparing the Dirichlet coefficients on both sides of the identity

$$\frac{\zeta'}{\zeta} + F = \left(\frac{\zeta'}{\zeta} + F\right)(1 - \zeta M) + \zeta' M + \zeta F M$$

gives for $n > v$:

$$-\Lambda(n) = -\sum_{\substack{k\ell=n \\ k>v,\ell>u}} \Lambda(k) \sum_{\substack{d|\ell \\ d>u}} \mu(d) - \sum_{\substack{k\ell=n \\ \ell\le u}} \log k\ \mu(\ell) + \sum_{\substack{k\ell m=n \\ \ell\le v, m\le u}} 1\cdot\Lambda(\ell)\mu(m)$$

This formula is originally due to Vaughan [Vau77]. Assume for now that $v \le N$ (we will eventually set u and v to be slightly less than $\sqrt{N}$). Multiply the above identity by $e(h\,\mathrm{Li}(n))$ and sum over n:

$$\begin{aligned}
\sum_{N<n\le N_2\le 2N} \Lambda(n) e(h\,\mathrm{Li}(n)) &= \sum_{u<\ell\le N_2/v} \sum_{\substack{N/\ell\le k\le N_2/\ell \\ v<k}} \Lambda(k) a(\ell) e(h\,\mathrm{Li}(k\ell)) \\
&\quad + \sum_{\ell\le u} \sum_{N/\ell\le k\le N_2/\ell} \mu(\ell)\log k\ e(h\,\mathrm{Li}(k\ell)) \\
&\quad - \sum_{r\le uv} \sum_{N/r\le k\le N_2/r} b(r) e(h\,\mathrm{Li}(kr)) \\
&= S_1 + S_2 - S_3,
\end{aligned}$$

where

$$a(\ell) = \sum_{\substack{d|\ell \\ d>u}} \mu(d), \text{ and } b(r) = \sum_{\substack{\ell m=r \\ \ell\le v, m\le u}} \Lambda(\ell)\mu(m).$$

Notice that $|a(\ell)|$ is at most $d(\ell)$, the number of divisors of ℓ, and similarly $|b(r)| \le \sum_{d|r} \Lambda(d) = \log r$, so we have the estimates

$$\sum_{L<\ell\le 2L} |a(\ell)|^2 \ll L\log^3 L, \text{ and } \sum_{R<r\le 2R} |b(r)|^2 \ll R\log^2 R.$$

It now suffices to show that $\sum_{0<h<H} |S_i| \ll N^{21/22+\epsilon}$ for each $i = 1, 2, 3$ by choosing u and v appropriately. We treat the sums of S_i individually in the next three subsections.

3.1. **The sum S_2.** Let $G(x) := \sum_{k\le x} e(h\,\mathrm{Li}(k\ell))$. By Lemma A.3, $G(x) \ll (xh\ell)^{\frac{1}{2}}\log(x\ell)$, so by partial integration we get

$$\begin{aligned} S_2 &= \sum_{\ell\le u}\mu(\ell)\sum_{N/\ell\le k\le N_2/\ell}\log k\ e(h\,\mathrm{Li}(k\ell)) \\ &\ll \sum_{\ell\le u}\left|\int_{N/\ell}^{N_2/\ell}\log x\ dG(x)\right| \\ &\ll \sum_{\ell\le u}\left(\sqrt{Nh}\log^2 N + \int_{N/\ell}^{N_2/\ell}\frac{1}{x}\sqrt{xh\ell}\log(x\ell)dx\right) \\ &\ll \sqrt{Nhu}\log^2 N. \end{aligned}$$

Thus $\sum_{1\le h<H}|S_2| \ll N^{21/22+\epsilon}$ (as desired) on taking $u = N^{5/11}$ and recalling that $H = \log^2 N$.

3.2. **The sum S_1.** Rewrite S_1 and split it into $\ll \log^2 N$ sums of the form:

$$\begin{aligned} S_1 &= \sum_{\substack{N\le k\ell\le N_2\\ v<k,u<\ell}}\alpha(k)\beta(\ell)e(h\,\mathrm{Li}(k\ell)) \\ &\ll \log^2 N\sum_{L<\ell\le 2L}\ \sum_{\substack{K<k\le 2K\\ N<k\ell\le N_2}}\alpha(k)\beta(\ell)e(h\,\mathrm{Li}(k\ell)). \end{aligned}$$

The roles of k and ℓ are essentially symmetric (allowing α and β to be either Λ or a affects only powers of log and not the final estimate) and taking $v = u$, we may arrange it so $N^{5/11} \le K \le N^{1/2} \le L \le N^{6/11}$.

Now using Lemma A.5, we find that:

$$\begin{aligned} S_1 &\ll \log^2 N\left(KL^{5/6}h^{1/6}\log^2 L\ \log^2 K\right) \\ &\ll \log^6 N\left(N^{21/22}h^{1/6}\right). \end{aligned}$$

Thus $\sum_h |S_1| \ll N^{21/22+\epsilon}$ as desired.

3.3. **The sum S_3.** Recall S_3 and break it according to:

$$\begin{aligned} S_3 &= \sum_{r\le uv} b(r)\sum_{N/r\le k\le N_2/r} e(h\,\mathrm{Li}(kr)) \\ &= \sum_{r\le u} + \sum_{u<r\le uv} \\ &= S_4 + S_5. \end{aligned}$$

We treat S_4 exactly as S_2, getting $S_4 \ll (Nh)^{1/2} \log N(u \log u)$, which is clearly sufficiently small.

For S_5, the analysis is identical to that of S_1 and gives the same estimate, so we are done.

Appendix A. Type I and II Estimates

We require the following two well-known estimates due originally to Weyl [Wey21] and van der Corput [vdC21, vdC22]; see e.g. Theorem 2.2 and Lemma 2.5 of [GK91].

Lemma A.1 (van der Corput)**.** *Suppose f has two continuous derivatives and $\Delta \ll f'' \ll \Delta$ on $[N, 2N]$. Then*

$$\sum_{N<n\leq N_1\leq 2N} e(f(n)) \ll N\Delta^{1/2} + \Delta^{-1/2}.$$

This is proved by truncating Poisson summation, comparing the sum to the integral, and integrating by parts two times.

Lemma A.2 (Weyl, van der Corput)**.** *Let $z_k \in \mathbb{C}$ be any complex numbers, $k = K+1, \ldots, 2K$. Then for any $Q \leq K$,*

$$\left|\sum_{K<k\leq 2K} z_k\right|^2 \leq \frac{K+Q}{Q} \sum_{|q|<Q} (1 - \frac{|q|}{Q}) \sum_{K<k, k+q\leq 2K} z_k \bar{z}_{k+q}.$$

To prove this, shift the interval by q and average the contributions over $|q| < Q$.

A.1. **Estimating Type I Sums.** We use Lemma A.1 to prove

Lemma A.3. *For any integer $\ell \geq 1$,*

$$\sum_{N<n\leq N_1\leq 2N} e(h \operatorname{Li}(n\ell)) \ll \begin{cases} N & \text{if } h = 0 \\ (N|h|\ell)^{\frac{1}{2}} \log(N\ell) & \text{otherwise.} \end{cases}$$

Proof. Let Ξ denote the sum in question. The trivial estimate is $\Xi \ll N$. Assume without loss of generality $h > 0$. Apply Lemma A.1 with $f(n) = h \operatorname{Li}(n\ell)$, taking $\Delta = \frac{h\ell}{N \log^2(N\ell)}$. Thus

$$\Xi \ll N \left(\frac{h\ell}{N \log^2(N\ell)}\right)^{1/2} + \left(\frac{N \log^2(N\ell)}{h\ell}\right)^{1/2} \ll (Nh\ell)^{1/2} \log(N\ell),$$

so we are done. □

A.2. **Estimating Type II Sums.** We first require the following estimate.

Lemma A.4. *For fixed integers $K < k \leq 2K$ and $1 \leq q < Q$ (where K and Q are some parameters) define following expression*

$$S_0(q;k) = \sum_{L<\ell\leq 2L} e(h(\mathrm{Li}(\ell k) - \mathrm{Li}(\ell(k+q)))). \tag{A.1}$$

Then

$$S_0 \ll (Lhq)^{1/2}.$$

Proof. We again apply Lemma A.1, now taking $f(x) = h(\mathrm{Li}(xk) - \mathrm{Li}(x(k+q)))$. Then

$$f''(x) = h\left(\frac{-k}{x\log^2(xk)} + \frac{k+q}{x\log^2(x(k+q))}\right) = hq\frac{\log(k'x)-2}{x\log(k'x)},$$

for some $k' \in [k, k+q)$ by the Mean Value Theorem in k. Thus we can take $\Delta = \frac{hq}{L}$ and

$$S_0 \ll L\left(\frac{hq}{L}\right)^{1/2} + \left(\frac{L}{hq}\right)^{1/2} \ll (Lhq)^{1/2},$$

as desired. □

With this estimate in hand, we control Type II sums as follows.

Lemma A.5. *Let $\alpha(\ell)$ and $\beta(k)$ be sequences of complex numbers supported in $(L, 2L]$ and $(K, 2K]$, respectively, and suppose that*

$$\sum_\ell |\alpha(\ell)|^2 \ll L\log^{2A} L \quad \textit{and} \quad \sum_k |\beta(k)|^2 \ll K\log^{2B} K.$$

Then

$$\sum_{L<\ell\leq 2L}\sum_{K<k\leq 2K} \alpha(\ell)\beta(k)e(h\,\mathrm{Li}(\ell k)) \ll KL^{5/6}h^{1/6}\log^A L\ \log^B K.$$

Proof. Let S denote the sum on the left hand side. By Cauchy-Schwartz,

$$|S|^2 \ll \left(\sum_\ell |\alpha(\ell)|^2\right)\sum_\ell\left|\sum_k \beta(k)e(h\,\mathrm{Li}(\ell k))\right|^2.$$

Let $Q \leq K$ be a parameter to be chosen later. Using Lemma A.2 and the supposed estimates on the second moments of α and β, we get:

$$\begin{aligned} |S|^2 &\ll L\log^{2A} L\,\frac{K+Q}{Q}\sum_{|q|<Q}(1-\frac{|q|}{Q}) \\ &\quad\times\sum_{\ell}\sum_{K<k,k+q\leq 2K}\beta(k)\bar{\beta}(k+q)e(h(\mathrm{Li}(\ell k)-\mathrm{Li}(\ell(k+q)))) \\ &\ll L\log^{2A} L\,\frac{K}{Q}\sum_{1\leq|q|<Q}\sum_{k}|\beta(k)\bar{\beta}(k+q)||S_0(q;k)|+\frac{K^2L^2}{Q}\log^{2A} L\ \log^{2B} K, \end{aligned}$$

where S_0 is defined by (A.1).

Using Cauchy's inequality, $|x\bar{y}| \leq \frac{1}{2}(|x|^2+|y|^2)$, and the fact that $|S_0(q;k)| = |S_0(-q;k+q)|$, we get

$$|S|^2 \ll \frac{K^2L^2}{Q}\log^{2A} L\ \log^{2B} K + \frac{LK}{Q}\log^{2A} L\sum_{k}|\beta(k)|^2\sum_{1\leq q<Q}|S_0(q;k)|.$$

From Lemma A.4 we have the estimate:

$$\frac{1}{Q}\sum_{1\leq q<Q}|S_0(q;k)| \ll (LhQ)^{1/2},$$

so we finally see that

$$|S|^2 \ll \frac{K^2L^2}{Q}\log^{2A} L\ \log^{2B} K + L^{3/2}K^2\log^{2A} L\ \log^{2B} KQ^{1/2}h^{1/2}.$$

The choice $Q = \lfloor L^{1/3}h^{-1/3}\rfloor$ gives the desired result.

□

REFERENCES

[Che73] Jing Run Chen. On the representation of a larger even integer as the sum of a prime and the product of at most two primes. *Sci. Sinica*, 16:157–176, 1973.

[FI98] John Friedlander and Henryk Iwaniec. The polynomial X^2+Y^4 captures its primes. *Ann. of Math. (2)*, 148(3):945–1040, 1998.

[GK91] S.W. Graham and G. Kolesnik. *Van der Corput's Method of Exponential Sums*, volume 126. London Math. Soc., Lecture Notes, 1991.

[HB01] D. R. Heath-Brown. Primes represented by x^3+2y^3. *Acta Math.*, 186(1):1–84, 2001.

[Lei77] D. Leitmann. The distribution of prime numbers in sequences of the form $[f(n)]$. *Proc. London Math. Soc.*, 35(3):448–462, 1977.

[Pu53] I. I. Pjateckiĭ-Šapiro. On the distribution of prime numbers in sequences of the form $[f(n)]$. *Mat. Sb.*, 33:559–566, 1953.

[Vau77] Robert-C. Vaughan. Sommes trigonométriques sur les nombres premiers. *C. R. Acad. Sci. Paris Sér. A-B*, 285(16):A981–A983, 1977.

[vdC21] J. G. van der Corput. Zahlentheoretische Abschätzungen. *Math. Ann.*, 84(1-2):53–79, 1921.

[vdC22] J. G. van der Corput. Verschärfung der Abschätzung beim Teilerproblem. *Math. Ann.*, 87(1-2):39–65, 1922.

[Vin37] I. M. Vinogradov. Representation of an odd number as a sum of three primes. *Dokl. Akad. Nauk SSSR*, 15:291–294, 1937.

[Wey21] H. Weyl. Zur Abschätzung von $\zeta(1+it)$. *Math. Z.*, 10:88–101, 1921.

E-mail address: `alexk@math.brown.edu`

DEPARTMENT OF MATHEMATICS, COLUMBIA UNIVERSITY, NEW YORK, NY, 10027

Current address: Department of Mathematics, Brown University, Providence, RI, 02912