AN EXPLICIT PRIME GAP UNDER THE RIEMANN HYPOTHESIS, FORMALIZED IN LEAN WITH ITS PRIME POWER INPUT MADE ELEMENTARY

STEVE HUBBARD

Abstract. We report on a Lean 4 formalization of Theorem 2 of Dudek, 2015, which states that if the Riemann hypothesis is true, then for every real $x \geq 2$ there is a prime in the interval $(x - (4/\pi)\sqrt{x}\log x, x]$. In the formalization, Theorem 2 is proved assuming the Riemann hypothesis and one cited explicit zero-counting bound of Hasanalizade, Shen and Wong, arXiv:2107.06506. Two inputs in the formal proof differ from those printed in Dudek's paper: first, the prime power contribution $\psi - \theta$ is required only across the moving window $(x-h, x+h]$ and is proved outright in Lean by an elementary counting argument with no cited input and without assuming the Riemann hypothesis; second, the sum over nontrivial zeros is bounded from the cited Hasanalizade, Shen and Wong zero count for $T \geq 70$ by trading slack between the head and tail sums at the split point $T = 2x/h$. The Riemann hypothesis is open and this note makes no progress on it.

1. The paper

This note documents a formalization in Lean 4 of the main explicit prime gap theorem from A. W. Dudek, *On the Riemann hypothesis and the difference between primes*, Int. J. Number Theory 11, 2015, no. 3, 771–778, doi:10.1142/S1793042115500426; preprint arXiv:1402.6417, v2, 20 May 2014 [Dudek]. That paper improves an earlier result of Ramaré and Saouter [RaSa].

The primary focus of this note is Theorem 2 of [Dudek]:

Theorem 1.1 (Dudek [Dudek], Theorem 2)**.** *Suppose the Riemann hypothesis is true. Then there is a prime in the interval*

$$\left(x - \frac{4}{\pi}\sqrt{x}\log x,\ x\right]$$

for all $x \geq 2$.

In addition to Theorem 2, Dudek's Theorem 1 gives a constant $c > 0$ such that

$$\pi(x + c\sqrt{x}\log x) - \pi(x) > \sqrt{x}$$

for all $x \geq 2$, and Dudek's Theorem 3 gives a prime in the interval

$$\left(x - (1+\varepsilon)\sqrt{x}\log x,\ x\right]$$

for all sufficiently large x. The same Lean entry also proves Theorems 1 and 3 under the same two hypotheses, with $c = 10^8$ for Theorem 1, whereas the remark of the paper that $c = 3 + \varepsilon$ works for large x is not part of the entry.

The purpose of this note is to record which inputs the formal proof took and how they differ from those printed in [Dudek]. No discrepancy was found in the

steps of Dudek's argument that were formalized. Theorem 2 is proved assuming the Riemann hypothesis and one cited zero count. The Riemann hypothesis is open and this note makes no progress on it. The binary Goldbach conjecture also remains open and is not treated here.

Outline of the paper's proof. Section 2 of [Dudek] establishes Theorem 2 using a weight w on the interval $(x-h, x+h]$ together with the integrated Chebyshev function

$$\psi_1(x) = \int_0^x \psi(t)\,dt.$$

The argument in Section 2 of [Dudek] proceeds through the following steps:

(1) **Explicit formula for ψ_1, Lemma 4 of** [Dudek]**.** For $x > 0$, with x not an integer,

$$\psi_1(x) = \frac{x^2}{2} - \sum_\rho \frac{x^{\rho+1}}{\rho(\rho+1)} - x\log(2\pi) + \varepsilon(x), \qquad |\varepsilon(x)| < \frac{12}{5},$$

where the sum runs over the nontrivial zeros ρ of the Riemann zeta function.

(2) **Weighted prime-power sum via second differences, Lemma 5 of** [Dudek]**.** For $x > 0$ and $h > 0$,

$$\sum_n \Lambda(n)w(n) = h - \frac{1}{h}\Sigma + \varepsilon(h), \qquad |\varepsilon(h)| < \frac{48}{5h},$$

where Λ is the von Mangoldt function and

$$\Sigma = \sum_\rho \frac{(x+h)^{\rho+1} - 2x^{\rho+1} + (x-h)^{\rho+1}}{\rho(\rho+1)}.$$

(3) **Splitting the sum over zeros.** Writing $\rho = 1/2 + i\gamma$ under the Riemann hypothesis, the sum Σ over nontrivial zeros is split at $|\gamma| = \alpha x/h$. To bound the head of the sum, the paper uses a zero count, specifically

$$N(T) < \frac{T\log T}{2\pi} \qquad \text{for } T > 15,$$

cited as Corollary 1 of Trudgian, Math. Comp. 81, 2012 [Tr].

(4) **Passing from Λ to primes.** To pass from the von Mangoldt function Λ to primes, the paper uses, in the paper's words, Theorem 14 and equation 5.5 of its reference [7], Rosser and Schoenfeld, 1962 [RS], namely the two-sided bound

$$0.98\sqrt{x} < \psi(x) - \theta(x) < 1.11\sqrt{x} + 3x^{1/3} \qquad \text{for } x \geq 121.$$

(5) **Positivity of the lower bound for $x \geq 65000$ and finite verification.** Choosing $\alpha = 2$ and

$$h = \frac{2}{\pi}\sqrt{x}\log x,$$

the last display of Section 2.2 of [Dudek] states that $\sum_{x-h<p\leq x+h} \log p$, the sum of $\log p$ over primes p in $(x-h, x+h]$, exceeds

$$\begin{aligned}\mathrm{rhs}(x) = h - \frac{2}{\pi}\left(\frac{x}{\sqrt{x-h}} + \frac{(x+h)^{3/2}}{x}\right)\log\left(\frac{2x}{h}\right) \\ - 1.11\sqrt{x+h} - 3(x+h)^{1/3} + 0.98\sqrt{x-h} - \frac{48}{5h},\end{aligned}$$

which is positive for all $x \geq 65000$. Positivity of $\mathrm{rhs}(x)$ guarantees the existence of a prime in $(x-h, x+h]$. Replacing $x+h$ by x then yields Theorem 2 from about 66798.7, namely

$$65000 + \frac{2}{\pi}\sqrt{65000}\log 65000.$$

The remaining values of x below this threshold were checked by computer in the paper, using Mathematica, with no table printed.

2. What the formal proof does

In the Lean 4 formalization [Palomar, Repo], the main theorem corresponding to Theorem 2 of [Dudek] has the full name

`LeanLab.PNTPort.ExplicitPsi1.Trivial.theorem2_of_RH_hsw`.

Its exact Lean declaration, together with the interval predicate and width function, is as follows:

```
theorem theorem2_of_RH_hsw (hRH : RiemannHypothesis)
    (hN : LeanLab.Dudek2014.SigmaH.HSWN70 gZ) :
    ∀ x : ℝ, 2 ≤ x → LeanLab.Dudek2014.PrimeInDudekInterval x
```

with

```
def PrimeInDudekInterval (x : ℝ) : Prop :=
  ∃ p : ℕ, p.Prime ∧ x - dudekH x < p ∧ (p : ℝ) ≤ x
noncomputable def dudekH (x : ℝ) : ℝ := 4 / π * √x * Real.log x
```

The hypotheses and objects appearing in this statement are defined as follows:

- `RiemannHypothesis` is the standard statement of the Riemann hypothesis from Mathlib [Mathlib].
- `gZ` is the family of imaginary parts of the zeros of `riemannZeta` with real part at least $-1/2$, which are the nontrivial zeros, indexed with multiplicity given by `analyticOrderNatAt`.
- `HSWN70 gZ` is the one cited input, included explicitly as a hypothesis to the theorem and never declared as an axiom.

The predicate `HSWN70` is defined in Lean by:

```
def HSWN70 (γ : ι → ℝ) : Prop :=
  (∀ T : ℝ, 70 ≤ T → ∀ s : Finset ι,
    ((s.filter (fun i => |γ i| ≤ T)).card : ℝ) ≤ 2 * (Mt T + Rt T)) ∧
  (∀ T t : ℝ, 70 ≤ T → T ≤ t → ∀ s : Finset ι,
    ((s.filter (fun i => T < |γ i| ∧ |γ i| ≤ t)).card : ℝ) ≤
      2 * ((Mt t + Rt t) - (Mt T - Rt T)))
```

where the main term $\mathrm{Mt}(T)$ and remainder bound $\mathrm{Rt}(T)$ are given by

$$\mathrm{Mt}(T) = \frac{T}{2\pi}\log\left(\frac{T}{2\pi e}\right) + \frac{7}{8},$$
$$\mathrm{Rt}(T) = 0.1038\log T + 0.2573\log\log T + 10.2425.$$

This hypothesis encodes Corollary 1.2 of Hasanalizade, Shen and Wong, arXiv:2107.06506 [HSW], which states that

$$\left|N(T) - \frac{T}{2\pi}\log\left(\frac{T}{2\pi e}\right)\right| \leq 0.1038\log T + 0.2573\log\log T + 9.3675 \qquad \text{for } T \geq e.$$

In the Lean formalization, this bound is used only for $T \geq 70$, with the constant $7/8$ moved into the error term so that

$$9.3675 + \frac{7}{8} = 10.2425,$$

and multiplied by 2 to count ordinates γ of both signs, $|\gamma| \leq T$ and $T < |\gamma| \leq t$. Consequently, the hypothesis `HSWN70 gZ` holds for the actual nontrivial zeros of the Riemann zeta function given Corollary 1.2 of [HSW], ensuring that `theorem2_of_RH_hsw` is not vacuous. The cited zero count itself is not proved in Lean.

TABLE 1. Comparison between the inputs and verification steps printed in Dudek [Dudek] and those used in the Lean 4 formalization.

Component	Printed paper [Dudek]	Lean 4 formalization
Zero hypothesis	Riemann hypothesis	`RiemannHypothesis` (Mathlib)
Zero count input	Trudgian [Tr], $T > 15$	`HSWN70 gZ` from [HSW], $T \geq 70$
Prime power step	Cited from [RS] for $x \geq 121$	`window_holds`, proved outright
ψ_1 error bound	$\lvert\varepsilon(x)\rvert < 12/5$	`error_bound` $\leq 2.393 < 12/5$
Large range	$x \geq 65000$ ($X \gtrsim 66798.7$)	`rhs_pos` ($x \geq 65000$), $X \geq 66799$
Small range	Mathematica check (unprinted)	`small_range` ($2 \leq x < 67001$, kernel)

Table 1 summarizes the relationship between the printed proof in [Dudek] and the formal proof. Two inputs differ from those printed in the paper: the prime power step and the zero sum estimate. We present the mathematics of each in the following two subsections.

2.1. **The prime power step, proved with no cited input.** A key observation of the formalization is that the proof of Theorem 2 uses the difference $\psi - \theta$ only through its increment across the window $(x - h, x + h]$. In Lean, this windowed inequality is named `SchoenfeldWindow` and is proved as the theorem `LeanLab.PrimePowerWindow.window_holds`:

```
def SchoenfeldWindow : Prop :=
  ∀ x : ℝ, 65000 ≤ x →
    (psiF (x + hD x) - thetaF (x + hD x)) -
      (psiF (x - hD x) - thetaF (x - hD x)) <
      1.11 * √(x + hD x) + 3 * (x + hD x) ^ (1 / 3 : ℝ) -
        0.98 * √(x - hD x)
```

where

$$\mathrm{hD}(x) = \frac{2}{\pi}\sqrt{x}\log x.$$

The theorem `window_holds` is proved outright in Lean, assuming no Riemann hypothesis and using no cited input. A companion lemma, `schoenfeld_window_of`, records that the two-sided bound of Rosser and Schoenfeld [RS] cited in the paper implies `SchoenfeldWindow`; thus the window form is a logical weakening of the printed input and is all that Dudek's argument requires.

Proposition 2.1 (`LeanLab.PrimePowerWindow.window_holds`)**.** *For every real $x \geq 65000$, with $h = \mathrm{hD}(x) = (2/\pi)\sqrt{x}\log x$,*

$$\big(\psi(x+h)-\theta(x+h)\big)-\big(\psi(x-h)-\theta(x-h)\big) < 1.11\sqrt{x+h}+3(x+h)^{1/3}-0.98\sqrt{x-h}.$$

Proof. Since $\psi(y) = \sum_{n\leq y}\Lambda(n)$ and $\theta(y) = \sum_{p\leq y}\log p$, the difference $\psi(y)-\theta(y)$ is the sum of $\Lambda(n)$ over non-prime integers $n \leq y$. Consequently, the window difference

$$\big(\psi(x+h)-\theta(x+h)\big)-\big(\psi(x-h)-\theta(x-h)\big)$$

is precisely the von Mangoldt mass of the prime powers p^k with $k \geq 2$ lying in the interval $(x-h, x+h]$.

First, the lemma `sum_nonprime_le` bounds the non-prime von Mangoldt mass over a natural-number interval $(A, B]$ with $A \leq B$:

$$\sum_{\substack{A<n\leq B\\ n \text{ not prime}}}\Lambda(n) \leq \sum_{2\leq k\leq \log_2 B}\frac{\log B}{k}\cdot\#\{m\in\mathbb{N} : A < m^k \leq B\}.$$

Indeed, every non-prime $n \in (A, B]$ with $\Lambda(n) \neq 0$ is of the form $n = p^k$ for a prime p and an integer $k \geq 2$ satisfying $k \leq \log_2 B$. For such n, its von Mangoldt weight satisfies

$$\Lambda(n) = \log p = \frac{\log(p^k)}{k} \leq \frac{\log B}{k},$$

and for each fixed exponent k, the map $n \mapsto p$ is injective.

Second, the lemma `count_le` bounds the number of k-th powers in the integer interval $(\lfloor a\rfloor, \lfloor b\rfloor]$ for real numbers $0 \leq a \leq b$:

$$\#\{m\in\mathbb{N} : \lfloor a\rfloor < m^k \leq \lfloor b\rfloor\} \leq b^{1/k} - a^{1/k} + 1.$$

Applying these two bounds to the window $(x-h, x+h]$ for $x \geq 65000$ and $h = \mathrm{hD}(x)$ separates the prime powers into squares, $k = 2$, and higher powers, $3 \leq k \leq \log_2 x$:

- For $k = 2$, there are at most $\sqrt{x+h}-\sqrt{x-h}+1$ squares in $(x-h, x+h]$, each carrying weight $\log p \leq \frac{1}{2}\log(x+h)$.
- For each $3 \leq k \leq \log_2 x$, there is at most one k-th power in $(x-h, x+h]$, carrying weight $\log p \leq \frac{1}{k}\log(x+h)$.

Setting $u = x^{1/6}$, the comparison between the resulting upper bound and the target expression $1.11\sqrt{x+h}+3(x+h)^{1/3}-0.98\sqrt{x-h}$ reduces to a polynomial inequality in u checked in Lean, using the four explicit estimates valid for all $x \geq 65000$:

$$u \geq 6.34, \qquad \log x \leq 1.5u+2.32, \qquad \log x \geq 11, \qquad \mathrm{hD}(x) \leq 0.63662\,u^3\log x. \quad \square$$

2.2. **The zero sum, from the Hasanalizade, Shen and Wong count.** The bound on the zero sum Σ is formalized as `LeanLab.Dudek2014.SigmaH.sigmaBd_of_hsw`, with conclusion `SigmaBd`, which is the bound of Section 2.2 of [Dudek] at $\alpha = 2$, on page 5 of the arXiv version:

Proposition 2.2 (`LeanLab.Dudek2014.SigmaH.sigmaBd_of_hsw`)**.** *Assume the Riemann hypothesis and `HSWN70 gZ`. For every $x \geq 65000$, with $h = \mathrm{hD}(x)$,*

$$|\Sigma| \leq \left(\frac{2xh}{\pi\sqrt{x-h}} + \frac{4h(x+h)^{3/2}}{2\pi x}\right)\log\left(\frac{2x}{h}\right).$$

The final bound has the same form as the bound in Dudek's paper. The route differs. From the count `HSWN70`, with its constant 10.2425, the separate head and tail estimates that would give this form do not both hold near $T = 70$: the tail inequality

$$\sum_{|\gamma|>T} \frac{1}{\gamma^2} \le \frac{\log T}{\pi T}$$

needs about $T \ge 120$, while at the split point $T = 2x/h$ with $x \ge 65000$ we only know $T \ge 70$.

The formal proof keeps the split at $T = 2x/h$ and trades slack between head and tail through three lemmas:

(1) **Head slack,** `head_count`**.** For all $T \ge 70$, the zero count hypothesis `HSWN70` implies

$$\#\{|\gamma| \le T\} \le \frac{T \log T}{\pi} - \big(0.89\,T - 22.25 - 0.502 \log T\big),$$

so the head count has a positive slack of about $0.89\,T$ below the main term $(T \log T)/\pi$.

(2) **Comparison of head and tail weights,** `weight_ratio`**.** Let

$$a = h^2(x-h)^{-1/2} \qquad \text{and} \qquad B = 4(x+h)^{3/2}$$

denote the head weight and tail weight, respectively. If $35h \le x$, which is equivalent to $T = 2x/h \ge 70$, then

$$B \le 1.1\, a\, T^2.$$

(3) **Combined split bound,** `split_bound`**.** Let D_i denote the second-difference term

$$D_i = \frac{(x+h)^{\rho_i+1} - 2x^{\rho_i+1} + (x-h)^{\rho_i+1}}{\rho_i(\rho_i+1)}$$

for the i-th zero. Using $B \le 1.1\, a\, T^2$ to absorb the excess from the tail sum into the $0.89\,T$ slack of the head count yields

$$\sum_i \|D_i\| \le \frac{T \log T}{\pi}\, a + \frac{B \log T}{\pi T}.$$

Substituting $T = 2x/h$, $a = h^2(x-h)^{-1/2}$, and $B = 4(x+h)^{3/2}$, this upper bound equals the displayed formula in `SigmaBd`.

2.3. **The rest of the formal proof.** The remaining steps of the formalization connect the explicit formula, the positivity of $\mathrm{rhs}(x)$, and the finite prime gap verification to assemble Theorem 2 for all $x \ge 2$:

- **Lemma 4 and Lemma 5,** `lemma4_Zpatch`**.** Lemma 4 of [Dudek] is proved in Lean as `lemma4_Zpatch`, combining the explicit formula `psi1_formula` and the error estimate `error_bound` for all $y > 0$ in the form

$$\left|\psi_1(y) - \left(\frac{y^2}{2} - Z(y) - y\log(2\pi)\right)\right| \le 2.393,$$

which is strictly below the printed bound $12/5 = 2.4$. Here $Z(y)$ is the zero sum for $y > 1$; for $y \le 1$ the Lean definition `Zpatch` takes the value that makes the identity exact, and the proof of Theorem 2 uses Z only for

$y > 1$. Lemma 5 of [Dudek] then follows by taking the second difference across $(x-h, x, x+h)$, giving the error bound $48/(5h)$.

- **Positivity for $x \geq 65000$, `rhs_pos`.** The lemma `rhs_pos` proves by explicit analytic bounds in Lean that $\mathrm{rhs}(x) > 0$ for every real $x \geq 65000$.
- **Large range transfer, `theorem2_large_of`.** From the existence of a prime in $(x-\mathrm{hD}(x), x+\mathrm{hD}(x)]$ for all $x \geq 65000$, the theorem `theorem2_large_of` deduces `PrimeInDudekInterval X` for all $X \geq 66799$ via the intermediate value theorem and monotonicity, since $65000 + (2/\pi)\sqrt{65000}\log 65000$ is about 66798.7.
- **Small range verification, `small_range`.** The theorem `small_range` proves `PrimeInDudekInterval x` for every real x satisfying $2 \leq x < 67001$, with no Riemann hypothesis, by a kernel prime gap scan using `decide +kernel` and no `native_decide`. On the main tier, `gaps_big` verifies that every integer $200 \leq n \leq 67000$ has a prime $p \leq n$ with $n - p < 72$. Five smaller tiers cover $2 \leq x < 200$ using the gap bounds $1, 2, 2, 4, 14$ starting from $x = 2, 3, 5, 7, 23$. Because the small range $[2, 67001)$ and the large range $[66799, \infty)$ overlap, `theorem2_of_RH_hsw` covers all $x \geq 2$.

3. Verification and disclosure

Verification artifacts.

- **Registry entry:** Palomar registry `PALOMAR-2026-11-01-000028`, version 1, `https://palomar-registry.org/entry.html?id=PALOMAR-2026-11-01-000028&version=1` [Palomar].
- **Repository:** `https://github.com/sthubbar/goldbach-lean`, directory `entries/dudek2014`, commit `da5cf9edad7896290771dfcab8feee2906399e3d` [Repo].
- **Toolchain:** Lean and Mathlib version `v4.35.0-rc3` [Mathlib].
- **Axioms and comparator check:** The axioms of the compared theorems are solely the standard Lean foundational axioms `propext`, `Quot.sound`, and `Classical.choice`, with `0 sorry`. The comparator passed against the challenge file `ChallengeDudek2014.lean`.

Disclosure. Steve Hubbard directed the project and decided what to publish. An AI agent, Claude by Anthropic, wrote the lab Lean code; sixteen ported files under `LeanLab/PNTPort/StrongPNT` are from the `PrimeNumberTheoremAnd` project [PNT], commit `c39a751132c88b6e8080b74c74023fd95b3d8be0`, Apache-2.0, credited to its contributors. Gemini, by Google, drafted the text of this note from a brief written by Claude under Steve Hubbard's direction, and Claude checked its statements against the Lean files. No human expert has reviewed the mathematics. The author of the paper has not been contacted.

References

[Dudek] A. W. Dudek, *On the Riemann hypothesis and the difference between primes*, Int. J. Number Theory 11 (2015), no. 3, 771–778, doi:10.1142/S1793042115500426, arXiv:1402.6417.

[HSW] E. Hasanalizade, Q. Shen and P.-J. Wong, *Counting zeros of the Riemann zeta function*, J. Number Theory 235 (2022), 219–241, arXiv:2107.06506.

[RS] J. B. Rosser and L. Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois J. Math. 6, 1962, no. 1, 64–94.

[RaSa] O. Ramaré and Y. Saouter, *Short effective intervals containing primes*, J. Number Theory 98, 2003, no. 1, 10–33.

[Tr] T. S. Trudgian, *An improved upper bound for the argument of the Riemann zeta-function on the critical line*, Math. Comp. 81, 2012, no. 278, 1053–1061.

[Palomar] Palomar registry, entry `PALOMAR-2026-11-01-000028`, version 1, `https://palomar-registry.org/entry.html?id=PALOMAR-2026-11-01-000028&version=1`.

[Repo] S. Hubbard, `goldbach-lean` repository, directory `entries/dudek2014`, commit `da5cf9edad7896290771dfcab8feee2906399e3d`, `https://github.com/sthubbar/goldbach-lean`.

[PNT] `PrimeNumberTheoremAnd`, `https://github.com/AlexKontorovich/PrimeNumberTheoremAnd`, commit `c39a751132c88b6e8080b74c74023fd95b3d8be0`.

[Mathlib] The mathlib Community, *The Lean mathematical library*, `https://github.com/leanprover-community/mathlib4`, version `v4.35.0-rc3`.

Independent Researcher