# Updating An Upper Bound Of Erik Westzynthius

Gerhard R. Paseman March 28, 2014

#### Abstract

Inspired by a paper of Erik Westzynthius, we build on work of Harlan Stevens and Hans-Joachim Kanold. Let k>2 be the number of distinct prime divisors of a positive integer n. In 1977, Stevens used Bonferroni inequalities to get an explicit upper bound on Jacobsthal's function g(n), which is related to the size of largest interval of consecutive integers none of which are coprime to n. Letting u(k) be the base 2 log of this bound, Stevens showed u(k) is  $O((\log k)^2)$ , improving upon Kanold's exponent  $O(\sqrt{k})$ . We use elementary methods similar to those of Stevens to get u(k) is  $O(\log k(\log \log k))$  in one form and  $O(\sigma^{-1}(n)\log k)$  in another form. We also show how these bounds can be improved for small k.

#### 1 Overview

Erik Westzynthius provided a ground-breaking result in [1] on prime gaps, showing that for any constant D > 0 there were infinitely many primes  $p_n$  so that  $p_{n+1} > p_n + D\log(p_n)$ . In the same paper, he also provided an upper bound for what was to be later called  $g(P_k)$ . This quantity measures how many consecutive integers we can find having a "small" ( $\leq p_k$ ) prime factor.

Ernst Jacobsthal in [2] defined and showed  $g(n) \leq k2^k + 2^k - k$ , where k is the number of distinct prime factors of n. Improved explicit bounds were later given by Kanold [3]  $(2^k$  for all k and  $2^{\sqrt{k}}$  for  $k \geq e^{50}$ ) and Stevens [4]  $(2k^{2+2e\log k})$ , and some additional but less explicit results given by Paul Erdős, Henryk Iwaniec, and others.

We show several explicit bounds for g(n), some depending on the quantities  $\sigma^{-1}(n)$  and  $\pi^{-1}(n)$ , including

$$g(n) < \frac{\sum_{i=1}^{s} {k \choose i}}{\pi^{-1}(n) - (\sigma^{-1}(n))^{s+1}/(s+1)!},$$

with s an odd integer bounded by (1 plus a constant times  $\sigma^{-1}(n)$ ), which gives  $g(n) < k^{B+C\log\log k}$  for k > 2 with explicit constants B and C, both < 3.9.

We define  $\sigma^{-1}(n)$  and  $\pi^{-1}(n)$  in the next section, and also list some results which apply when  $\sigma^{-1}(n)$  is small ( $< 1+1/2q_1$ ). The following section recalls work of Jacobsthal and Westzynthius, and shows how Stevens's bound can be tightened with a little effort.

With elementary means, we also show the bound above that uses  $\sigma^{-1}(n)$  and  $\pi^{-1}(n)$ , and follow that section with some supplementary results as well as suggestions for further research. The remainder of this article contains some history, a recommended reading list, and an Appendix as well as acknowledgments and a list of citations.

# 2 Definitions and simple bounds

We use  $\omega(n)$  for the number of distinct (positive) prime factors of the positive integer n, and declare  $k = \omega(n)$ . We also require k > 0, so n > 1. We list the prime factors  $q_i$  of n in increasing order:  $q_1 < q_2 < \ldots < q_k$ . We recall the kth primorial  $P_k$  as  $P_k = \prod_{1 \le i}^k p_i$ .

**Definition**: We define (here  $^{-1}$  is part of a label,  $\overline{not}$  an exponent)

$$\sigma^{-1}(n) = \sum_{1 \le i}^{k} 1/q_i$$
, and  $\pi^{-1}(n) = \prod_{1 \le i}^{k} (1 - 1/q_i)$ .

Note that  $n\pi^{-1}(n) = \phi(n)$ , Euler's function for counting positive integers coprime to and less than n.

**Definition**: After Jacobsthal, for n > 0 define g(n) to be the smallest positive integer m such that for any integer a, the set of m consecutive integers  $\{a+1,\ldots,a+m\}$  has an integer a+j, where  $1 \leq j \leq m$ , such that  $\gcd(a+j,n)=1$ . If n>1, define L(n) as the largest integer l so that there is an interval of l consecutive integers  $\{b+1,\ldots,b+l\}$  such that each b+j satisfies  $\gcd(b+j,n)>1$ .

It is straightforward to show g(n) = L(n) + 1 for all n > 1. Also  $g(n) = \max_i (c_{i+1} - c_i)$  where the  $c_i$  denote all integers coprime to

(totatives of) n in increasing order. Consequently, g(n) depends only on the set of distinct prime factors of n; we will use n squarefree at times in this article.

One has  $g(p_i) = 2$ , and g(n) < n for n > 2. If n has "large" prime factors  $(q_i > k \text{ for all } i)$ , then g(n) = L(n) + 1 is "small". Specifically, **Proposition**(Jacobsthal): If  $q_1 > k$ , then L(n) = k.

Proof Sketch: Any integer interval of length k+1 has at most one multiple of  $q_i$ , giving at most k integers in that interval having a prime factor in common with n. So at least one integer in that interval is not a multiple of any  $q_i$ . So L(n) < k+1.

Conversely, for any permutation  $\tau$  of the k indices, the Chinese Remainder Theorem gives an integer  $b_{\tau}$  such that  $b_{\tau} + \tau(i) = 0 \mod q_i$ for  $1 \leq i \leq k$ , so  $L(n) \geq k$ . **End** of Proof Sketch.

We could have chosen  $\tau$  to be the identity permutation, but we want to point out that this proof gives k! many intervals in (0,n) which achieve the L(n) bound for this kind of n. This proof also shows g(n) > k for any n > 1.

The next bound is inspired by Kanold's 1967 paper. We would like to know if it appears explicitly in the literature.

**Proposition**(Kanold-**P**.): Let 0 < r < 1 be a real number and n such that  $r + \sigma^{-1}(n) < 1$ . Then L(n) < k/r, and  $g(n) \le \lceil k/r \rceil$ .

Proof Sketch: For an interval containing L consecutive integers, at most  $\lceil L/q_i \rceil \leq 1 + (L-1)/q_i$  of them are multiples of  $q_i$ . Since  $\sigma^{-1}(n) < 1 - r$ , the count of numbers not coprime to n is at most  $(L-1)\sigma^{-1}(n) + k < (L-1)(1-r) + k$ . Now whenever  $L \geq k/r$ ,  $(L-1)(1-r) + k \leq L - 1 + r$ , so this count is less than L. One gets  $g(n) \leq \lceil k/r \rceil$ , giving L(n) < k/r. **End** of Proof Sketch.

This gives a weaker bound than Jacobsthal's proposition, but it applies to more cases, even when  $q_1$  is about  $2\sqrt{k}$  and  $r > 1/q_1$ . It suggests the following

**Variation**: Let k > 1 and  $\sigma^{-1}(n) < 1 + \frac{1}{2q_1}$ . Then

$$L(n) < \frac{q_1}{q_1 - 1} (2k - 1 - \sigma^{-1}(n)) 2q_1$$
, so  $g(n) < 4kq_1 \frac{q_1}{q_1 - 1}$ .

Proof Sketch: The estimate above for multiples of  $q_i$ ,  $1 < i \le k$ , among the L numbers is refined by subtracting that portion which are also multiples of  $q_1$ ; we underestimate it by  $L/q_iq_1 - 1 < \lfloor \lceil L/q_i \rceil/q_1 \rfloor$ . Thus  $(L-1)\sigma^{-1}(n) + k - \sum_{1 < i}^k (-1 + L/(q_iq_1))$  is the refined upper bound for the non-totative count. Rewriting, we look for L so that

$$L[\sigma^{-1}(n)(1-1/q_1)+1/(q_1)^2]+2k-1-\sigma^{-1}(n) < L,$$

which holds if and only if

$$L(1-1/q_1)(1+1/q_1-\sigma^{-1}(n)) > (2k-1-\sigma^{-1}(n)).$$

The assumption  $(1 + 1/q_1 - \sigma^{-1}(n)) > 1/2q_1$  yields that if  $L/2q_1 > (2k - 1 - \sigma^{-1}(n))(q_1/(q_1 - 1))$ , then L is greater than our overcount; as above, we get  $L(n) < 2q_1(2k - 1 - \sigma^{-1}(n))(q_1/(q_1 - 1))$  and the weaker  $g(n) < 4kq_1(q_1/(q_1 - 1))$ . **End** of Proof Sketch.

We now present one bound of general character, and one which asymptotically improves upon Kanold's smaller bound. We use the

**Fact**: Let b, d and f be integers with gcd(f, d) = 1. There is an integer z so that for all integers t, gcd(b + tf, d) = gcd(zb + t, d).

Pick z so that  $zf = 1 \mod d$ .  $\gcd(b + tf, d) = \gcd(z(b + tf), d) = \gcd(zb + t, d)$ , giving the Fact above and the application below that coprimality in sequences of consecutive integers behaves the same way in certain arithmetic progressions.

Assume k > 2, n squarefree and  $d \mid n$  with d neither 1 nor n. Set f = n/d. 1+tf is coprime to f, and the Fact above shows that 1+tf always has at least one out of g(d) consecutive members coprime to d. Any interval of (g(d)f)-many integers thus has one coprime to n of the form 1+tf. Considering c+tf for all totatives c of f, we get the

**Observation**:  $g(fd) \le g(d)f - f + g(f)$ .

Proof Sketch: When  $c \neq 1$ , there is at least one number of the form c+tf coprime to n=fd in the interval (1+a,1+a+g(d)f), where  $a=t_0f$  and we assume that if 1+tf is coprime to n, then it is outside this interval. There are at least  $(\phi(f)-1)$ -many of these numbers; try placing them inside the interval leaving a large gap. If you can get a gap of at least g(d)f-f, then the numbers must be in the subintervals (1+a,1+a+f) and (1+a+(g(d)-1)f,1+a+g(d)f). If the largest in the first interval is c+a, then the smallest in the second interval is at most c'+a+(g(d)-1)f, where c' is the next largest totative to f after c. But  $c'-c \leq g(f)$ , giving the bound. **End** of Proof Sketch.

We use weaker forms of the Observation and Variation to improve on Kanold's smaller bound.

Let n be squarefree. Consider the tail of  $\sigma^{-1}(n)$ , that is, find  $q_l$  smallest such that  $\sum_{l \leq i}^k 1/q_i < 1 + 1/2q_l$ . Let  $d = \prod_{l \leq i}^k q_i$ . From the Variation we have  $g(d) < 4(k-l+1)(q_l^2/(q_l-1))$ . However, when k is large,  $q_l$  is small enough that n/d is smaller than  $2^{\sqrt{k}}$ .

**Improvement**: For k sufficiently large,  $n/d < 2^{\frac{3}{2}k^{0.45}}$ , giving  $g(n) < 2^{3k^{0.45}/2} \frac{4(k-l+1)q_l^2}{q_l-1}$ . As k gets large,  $\log g(n) < k^{(\epsilon+1/e)}$  eventually for any fixed  $\epsilon > 0$ .

Proof Idea: Let m satisfy  $p_m = q_{l-1}$ . We defer to the Appendix showing that  $p_m < k^{0.45}$  when  $k \ge e^{9.5}$ . As  $n/d \le P_m$ , and for an explicit positive c < 0.02,  $P_m < e^{p_m(1+c)} < 2^{3p_m/2}$ ; the Observation implies the weaker form  $g(n) < (n/d)g(d) < 2^{3k^{0.45}/2}g(d)$ .

implies the weaker form  $g(n) < (n/d)g(d) < 2^{3k^{0.45}/2}g(d)$ . About  $\log g(n)$ , we use that  $\sum_{l \le j}^k 1/q_j \le \sum_{1 \le i}^{k-l+1} 1/p_{m+i}$ , and then truncate to the first  $u \le k-l+1$  terms while still having  $\sum_{1 \le i}^u 1/p_{m+i} > 1 - 1/2p_m$ . We use an approximation of Mertens for this last sum to get  $\log(\log(p_{m+u})/\log p_m) > 1 - O(1/\log p_m)$ . Picking a convenient positive  $\epsilon_1$  and  $\epsilon$ , as m goes large,  $p_{m+u} > p_m^{e-\epsilon_1}$  eventually, and  $\log$  of the bound is majorized by  $p_m(1+c')$  for some positive small c' and is eventually dominated by  $k^{1/e+\epsilon}$ . The Appendix provides more detail. **End** of Proof Idea.

We take credit for this form of presentation, but are influenced by Kanold's paper; his proof for  $2^{\sqrt{k}}$  has many of the ideas above, and we wonder if perhaps he did discover it.

We end this section with lower bounds: Westzynthius shows an easily demonstrated lower bound  $2p_{k-1} \leq g(P_k)$  for k > 1, and then shows a better lower bound (for k sufficiently large after choosing  $\epsilon > 0$ ) of  $e^{\gamma}(1 - \epsilon)p_k \frac{\log \log p_k}{\log \log \log p_k}$ .

#### 3 Improving estimates of Stevens

Both Westzynthius and Jacobsthal use a simple sieve argument to establish an upper bound on g(n), using n squarefree. Since Stevens uses part of this, we show the argument here.

Recap: We use inclusion-exclusion to count integers in the open interval L=(a,a+x) that are coprime to squarefree n, for real numbers a and x>0.  $I_0$  counts the totatives of n in L, and for t>0 and  $t\mid n$ ,  $I_t$  counts multiples of t in L. Then  $I_0=\sum_{t\mid n}(-1)^{\omega(t)}I_t$ . As in the previous section we replace the count  $I_t$ , this time by x/t+E(t). E(t) is an error term which depends on both t and a actually, but always  $|E(t)| \leq 1$ , and will be removed below.

$$I_0 = \sum_{t|n} (x/t + E(t))(-1)^{\omega(t)} \ge x \sum_{t|n} (-1)^{\omega(t)}/t - \sum_{t|n} 1.$$

Now rewrite the sum of  $(-1)^{\omega(t)}/t$  as a product  $\prod (1-1/q_i)$ , and note the second sum is  $2^k$ . We get

$$I_0 \ge x\pi^{-1}(n) - 2^k.$$

Now this relation above is essentially independent of a. If we pick x large enough, then  $I_0 > 0$  and g(n) will be at most x. So choose  $x = 2^k/\pi^{-1}(n) + \epsilon$ . **End** of Recap.

The above shows  $g(n) \leq 2^k/\pi^{-1}(n)$ . One would like to improve on this since Kanold has an asymptotically better result.

In our view Stevens has two main ideas in his 1977 paper. The first is to observe that the sum produced from inclusion-exclusion can be truncated a la Bonferroni to a smaller sum to give fewer terms to approximate. The second is that the denominator  $T_s$  can be written as  $\pi^{-1}(n) - T'_s$ , and that  $T'_s$  can be easier to handle. We adapt his proof slightly, and then we tighten up the estimates he uses to improve the exponent.

Adaptation: We use  $x, I_0, t$ , and  $I_t$  where Stevens used Q, L, B and N(...). We use integral x, following Stevens. We assume k > 4 to make sure some of his estimates apply. He reorganizes the sum by the number of factors in t and uses a result of Landau for the first idea. His display (3) in our notation says: for any odd value of s,

$$I_0 \ge \sum_{0 \le i}^{s} (-1)^i \sum_{t \mid n, \omega(t) = i} I_t.$$

Using the estimate  $|I_t - x/t| \le 1$  (except for t = 1 when  $I_1 = x$ ), we write what Stevens has in his (4) and (5) as

$$I_0 \ge x \sum_{0 \le i}^s (-1)^s \sum_{t \mid n, \omega(t) = i} \frac{1}{t} - \sum_{1 \le i}^s {k \choose i} = \det xT_s - SB.$$

(We've written SB for  $\sum_{1\leq i}^s {k \choose i}$ , and  $T_s$  for  $\sum_{0\leq i}^s (-1)^s \sum_{t\mid n,\omega(t)=i} \frac{1}{t}$ .) Defining  $T_s'$  by the relation  $\pi^{-1}(n) - T_s' = T_s$ , Stevens notes (using an approximation of  $k^s$  for SB instead of SB directly) that if s is chosen so that  $\pi^{-1}(n) > T_s'$  (so that  $T_s > 0$ ), and if  $x > SB/(\pi^{-1}(n) - T_s')$ , then  $I_0 > 0$  and so  $g(n) \leq x$ . He and we now look for a suitable s.

 $T'_s$  can be written as a sum over i of sums of terms t with  $\omega(t) = i$ , just like  $T_s$ , but with  $s < i \le k$ . Before doing this, Stevens observes: for k > 2 (we insert h(k), a putative upper bound for  $\sigma^{-1}(n)$ )

$$r! \sum_{t|n,\omega(t)=r} 1/t < (\sigma^{-1}(n))^r < h(k)^r.$$

Stevens uses  $\log k$  for h(k); later we will use  $\log \log p_k + 1/2$ . Then

$$T'_{s} = \sum_{s < r}^{k} (-1)^{r} \sum_{t \mid n, \omega(t) = r} 1/t < \sum_{s < r}^{k} \sigma^{-1}(n)^{r}/r! < \sum_{s < r}^{\infty} h(k)^{r}/r!$$

which follows by dropping the  $(-1)^r$  and by extending the sum past r = k. Taylor's theorem with remainder on  $e^{h(k)}$  then yields

$$T'_s < \sum_{s < r}^{\infty} h(k)^r / r! \le e^{h(k)} h(k)^{s+1} / (s+1)!.$$

Stevens bounds things further by asking  $s+1 \ge 2eh(k)$  and using  $(s+1)! > ((s+1)/e)^{s+1}$ . Then

$$T_s' < e^{h(k)}h(k)^{s+1}/(s+1)! < e^{h(k)}(eh(k)/s+1)^{s+1} \le e^{h(k)}/2^{s+1}.$$

He also under-estimates  $\pi^{-1}(n)$  by 1/k for k > 4, where we will use  $\beta/\log k$ . He then has

$$\pi^{-1}(n) - T'_s > 1/k - k/2^{s+1} > 1/k - k^{1-2e\log 2} > 0,$$

since s+1 is an even integer greater than  $2e\log k$ . So  $I_0>0$  (and thus  $g(n)\leq x$ ) if s is an odd integer greater than  $2e\log k-1$  and  $x>2k^{s+1}>k^s/(1/k-k/2^{s+1})>SB/(\pi^{-1}(n)-T_s')$ . Stevens replaces s+1 with  $2e\log k+2$  to ensure the bound holds for all k. **End** of Adaptation.

We repeat the above, using  $\log \log p_k + 1/2$  for h(k) and  $\beta/\log p_k$  for  $\pi^{-1}(n)$ . (The Appendix discusses the validity of these choices.) Again asking for odd s with  $s+1 \geq 2eh(k)$ , then  $T'_s \leq e^{h(k)}/2^{s+1}$ , and

$$\pi^{-1}(n) - T'_{s} > \beta/\log p_{k} - (\log p_{k})e^{1/2}/2^{s+1}$$

$$> \beta/\log p_{k} - (\log p_{k})e^{1/2}/2^{2e(\log\log p_{k}+1/2)}$$

$$= \beta/\log p_{k} - (\log p_{k})(e/2^{2e})^{1/2}/(\log p_{k})^{2e\log 2}$$

$$> (1/\log p_{k})[\beta - (\log p_{k})^{2(1-e\log 2)}/3] > 0.$$

In the last line, we use that  $(e/2^{2e})^{1/2} < 1/3$ , that we can pick  $1/1.78 > \beta > 1/3$ , that  $\log p_k > 2$  because k > 4, and that  $e \log 2 > 1$ .

It should be clear that  $\pi^{-1}(n) - T'_s > 1/4 \log p_k$  by choosing  $\beta > 1/3$  and whenever  $(\log p_k)^{2e \log 2 - 2} > 4$  which holds for k > 4. Using such an estimate we have whenever s + 1 is even and  $\geq 2e(\log \log p_k + 1/2)$ ,  $\pi^{-1}(n) - T'_s > 1/4 \log p_k$  leading to

**Theorem:** 
$$g(n) \le (4 \log p_k) \sum_{1 \le i \le s} {k \choose i} < (4 \log p_k) k^{1 + 2e(1/2 + \log \log p_k)}$$
.

We could tweak the choice of s slightly to get a smaller exponent, as well as use a better approximation for the sum of binomial coefficients. In the next section, we will find a bound which depends directly on  $\sigma^{-1}(n)$  which not only does both, but gives a tighter bound in general.

**4** 
$$\sigma^{-1}(n)$$
 and  $\pi^{-1}(n)$ 

We modify Stevens's argument with a better upper bound for the numerator and express  $T'_s$  as an alternating and eventually decreasing sum, allowing us a smaller s.

Note that  $T'_s$  is an alternating sum and that one has

$$\sigma^{-1}(n) \sum_{t \mid n, \omega(t) = j} 1/t > (j+1) \sum_{t \mid n, \omega(t) = j+1} 1/t,$$

so that when  $s > \sigma^{-1}(n)$ , one can bound  $T_s'$  by  $\sigma^{-1}(n)^{s+1}/(s+1)!$ . Now instead of Taylor's theorem and h(k), we use (see Appendix)  $e < (1/\pi^{-1}(n))^{1/\sigma^{-1}(n)} \le 4$  to show any real number  $r \ge 4\sigma^{-1}(n)$  gives  $\pi^{-1}(n) - T_{\lceil r \rceil - 1}' > \pi^{-1}(n) - \sigma^{-1}(n)^{\lceil r \rceil}/(\lceil r \rceil)! > 0$ :

$$e < 4^{3/4}, \text{ so } e(1/\pi^{-1}(n))^{1/4\sigma^{-1}(n)} < e4^{1/4} < 4$$
so  $\sigma^{-1}(n) < \pi^{-1}(n)^{1/4\sigma^{-1}(n)} 4\sigma^{-1}(n)/e \le \pi^{-1}(n)^{1/r}(r/e)$ 
so  $\sigma^{-1}(n)^{\lceil r \rceil} < \pi^{-1}(n)(\lceil r \rceil/e)^{\lceil r \rceil} < \pi^{-1}(n)(\lceil r \rceil)!$ .

We now claim

**Theorem:** Let s be the smallest odd integer with  $s+1 \ge 4\sigma^{-1}(n)$ . For k > 2,

$$\frac{\sum_{1 \le i \le s} {k \choose i}}{\pi^{-1}(n) - \sigma^{-1}(n)^{s+1}/(s+1)!} > g(n).$$

We collapse the summands in the numerator slightly, increasing the total by 1, and as  $(s+1)! > \sqrt{2\pi(s+1)}((s+1)/e)^{s+1}$ , one sees the denominator is larger than  $(\sqrt{2\pi(s+1)}-1)\sigma^{-1}(n)^{s+1}/(s+1)!$ , so we can write a weaker upper bound as a **corollary**:

$$\frac{(s+1)! \sum_{0 \le 2j < s} {k+1 \choose s-2j}}{(\sqrt{2\pi(s+1)} - 1)\sigma^{-1}(n)^{s+1}} > g(n).$$

This may seem intimidating, but when we take into account that for k>6,  $\sigma^{-1}(n)$  is at most  $1/2+\log\log k(1+\log 2/\log k)$  it is then seen that the dominant term in the sum is  $\binom{k+1}{s}$  and the expression is  $O(((k+1)/\sigma^{-1}(n))^{3+4\log\log k+\epsilon})$  when  $\sigma^{-1}(n)\geq 1$ . (The portion that is  $\frac{s+1}{(\sqrt{2\pi(s+1)}-1)\sigma^{-1}(n)}$  is less than 1 for large enough  $\sigma^{-1}(n)$ ; for  $\sigma^{-1}(n)$  near or smaller than 1 we have the more elementary bounds.) This argument only needs K such that  $(s+1)>K\sigma^{-1}(n)$  and also that  $e/K<\pi^{-1}(n)^{1/K\sigma^{-1}(n)}$ . This holds for K>3.89, and when

 $\sigma^{-1}(n) > 1$ , one can lower K from 4 to 3.81. However, even for large values of  $\sigma^{-1}(n)$ , the argument still expects  $(K/e)^K > e$ , which means K can't be shown smaller than 3.59 with this method.

As a rough comparison, Jacobsthal's bound is larger than Kanold's bound for all k. Stevens's bound is smaller than  $2^k$  for k>300, and is smaller than Kanold's better bound for k>5000000. Our Improvement is smaller than Kanold's for  $k>e^{11}$ , and the exponent  $s\leq 1+4\sigma^{-1}(n)$  above is smaller than that of Stevens for k>2.

### 5 Some History

The recap is our interpretation of Westzynthius's upper bound argument published in 1931, generalized to arbitrary n with  $\omega(n) = k$  instead of  $P_k$ , which Westzynthius did not publish as far as we know. In a footnote Westzynthius did hint at sieving with just odd numbers, and we considered extending that argument with thinner sets. This led us to asking the question [6] on MathOverflow in 2010.

Correspondence on MathOverflow led us eventually to Thomas Hagedorn's paper [5] and Jacobsthal's function. Jacobsthal in [2] uses a slightly different argument, and (with the notation of this article) instead of using  $2^k/\pi^{-1}(n)$  he bounds  $\pi^{-1}(n)$  by 1/(k+1) and gives a bound of  $(k+1)(2^k-1)$  on L(n). Hagedorn's paper quoted the bounds of Kanold and Stevens, and after studying those papers we adapted Stevens's argument and posted the results on MathOverflow in 2011, as well as producing a private manuscript with small circulation.

Since then we have accumulated and posted other accessible results, and arranged some of them for this article. The Observation represents a small improvement on Kanold's result which involves  $g(fd) \leq g(d)f + 1 - \phi(f)$  instead of  $g(fd) \leq g(d)f + g(f) - f$ ; the two are the same for f prime. The Improvement is intended to show not just improved asymptotic results but also that Kanold's bound holds for k smaller than  $e^{50}$ . Indeed the name is earned once  $k \geq e^{11}$ . We admit the work is in showing the bounds hold for small k, which makes the Improvement not as elementary or accessible as we hope.

Hagedorn also mentions work of Erdős, Iwaniec, and others. Erdős shows for any given positive real  $\epsilon$  that  $|1-g(n)\pi^{-1}(n)/k| > \epsilon$  only for n in a set of zero density. Erdős also comments that Brun's method can yield a constant c such that g(n) is  $O(k^c)$ , but we have not found a version of this that is both explicit and accessible. Iwaniec shows

the existence of a constant C independent of n such that there are at least  $k^2$  many totatives of n in an interval of size  $C(k^2 \log k)/\pi^{-1}(n)$ , which implies g(n) is  $O((k \log k)^2)$ ; again we do not know what C is.

#### 6 Further research and reading

We intended this article to give simpler, more accessible, and more explicit proofs of upper bounds on Jacobsthal's function. We are optimistic about improving upon the results of Erdős and Iwaniec. In particular, we think there is more to the Observation: we hope to achieve a subquadratic in k upper bound using this direction by noting how large intervals of numbers with factor common to n are distributed in (0,2n). At present, the difference in (base 2) exponents between  $O(\log k \log \log k)$  and  $O(\log k)$  is substantial.

Except for the bound depending on  $\sigma^{-1}(n)$ , all of these bounds are also bounds on Jacobsthal's C(k), given by  $C(k) = \max_{k=\omega(n)} L(n)$ . It was shown recently [8] that the conjecture  $C(k) + 1 = g(P_k)$  holds for  $1 \le k \le 23$  and fails at k = 24.

Note that the bound involving  $\sigma^{-1}(n)$  can represent a substantial improvement even if n cannot be factored; for those n which do not have small factors,  $\sigma^{-1}(n)$  can be substantially smaller as can k, even for numbers near  $10^{10^{100}}$ . Of course, when k or a partial factorization of n are better known, better bounds on g(n) become available.

We are interested in tweaking the Variation to handle more squarefree n by sieving out small prime factors. Our beginning efforts have not yielded much improvement on bounds obtained by the Observation. It seems better estimates on the number of totatives in an interval of arbitrary length are needed to carry out an argument like that in the Variation.

Some questions of interest:

- 1) Pick a small odd prime p and odd n with  $q_1 > p$ . We know g(pn)/g(n) < p: can we get anything sharper? In particular, what are those integers n such that g(3n) > 2g(n)? Such that g(5n) > 2g(n)?
- 2) Let a(n) be the smallest positive integer such that gcd(n, a(n) + i) > 1 for 0 < i < g(n). One can show a(n) < n/2; how much can this be improved? If b(n) is the number of such longest intervals of nontotatives of n in (1, n), can we hope for a(n)b(n) < n?
- 3) How close are two such intervals? If one hopes for a subquadratic (in k) bound on q(n), this will be an important bit of information.

Even in the case  $q_1 > k$ , it should be related to how primes are distributed, which suggests that some interesting perspective is needed.

- 4) Not much asymptotic improvement should be expected from these arguments in the case that  $\sigma^{-1}(n)$  is large, say  $\sigma^{-1}(n) > 2$ . However, that is where the difficult cases are, and the quantity  $\mathcal{T} = (1/\pi^{-1}(n))^{1/\sigma^{-1}(n)}$  is expected to decrease as  $\sigma^{-1}(n)$  increases. How does  $\mathcal{T}$  behave with  $\sigma^{-1}(n)$ , and can one use this in bounding g(n)?
- 5) Even the simple estimates with small  $\sigma^{-1}(n)$  have some slop, primarily in overestimating multiples of  $q_i$  with  $L/q_i$ . Often this results in an estimate about twice as large as needed. Can something be said about this "noise" vector  $L/q_i I_{q_i}$  and what approaches avoid the error introduced by this?

We recommend the bibliography and also the following reading list, which provides additional information related to Jacobsthal's function and applications.

Ernst Jacobsthal, Über Sequenzen ganzer Zahlen, von denen keine zu n teilerfremd ist. I-III. Det Kongelige Norske Videnskabers Selskabs Forhandlinger Bd 33 1960, Nr. 24, Trondheim I Kommisjon Hos F. Bruns Bokhandel 1961, pp. 117-124,125-131,132-139. (Also see IV and V published in a later edition.)

Hans-Joachim Kanold, Neuere Untersuchungen über die Jacobsthal-Funktion g(n). Monatshefte Math. 84, 1977, pp. 109-124.

Paul Erdős, On the integers relatively prime to n and on a number-theoretic function considered by Jacobsthal. *Math. Scand.* 11 (1962) pp. 163-170.

R.C. Vaughan, On the order of magnitude of Jacobsthal's function. *Proc. Edinburgh Math. Soc.* 20 (1976-77) pp. 329-331.

Henryk Iwaniec, On the problem Of Jacobsthal. *Demonstratio Mathematica* v XI no. 1 1978 pp.225-231.

MathOverflow Questions (Number refers to URL, so 37679 expands to http://mathoverflow.net/questions/37679)

88323 Analogues of Jacobsthal's Function.

56099 Lower bound of the number of relatively primes (each other) in an interval.

68351 Least Prime Factor in a sequence of 2n consecutive integers.

Shallit, J., State Complexity and Jacobsthal's Function. CIAA '00 Revised Papers from the 5th International Conference on Implementation and Application of Automata 2000, pp. 272-278

Schlage-Puchta, J.C., On Triangular Billiards. http://arxiv.org/abs/1105.1629

#### 7 Acknowledgments

We acknowledge Will Jagy, Aaron Meyerowitz, and Thomas Hagedorn for their support and assistance. We appreciate the stimulating environment provided by the MathOverflow forum and its community, and are thankful for its providing a repository of these results. We especially appreciate Will and Aaron for their MathOverflow contributions, and are also grateful for discussions with users Larry Freeman (who asked a version of question 2) above) and asterios gantzounis.

### 8 Appendix

We resolve some details on assertions made in earlier sections: that  $(1/\pi^{-1}(n))^{1/\sigma^{-1}(n)} \in (e, 4]$ , on bounding  $\sigma^{-1}(n)$  from above by  $1/2 + \log \log p_k$  and  $\pi^{-1}(n)$  from below by  $1/3 \log p_k$ , and on showing the Improvement holds for  $k > e^{9.5}$ .

In getting a bound depending on  $\sigma^{-1}(n)$ , we used the assumption that  $e < (1/\pi^{-1}(n))^{1/\sigma^{-1}(n)} \le 4$  for n > 1. We proved this along with related results in a private manuscript [7]. The proof was based on observing that  $(-\log \pi^{-1}(n))/\sigma^{-1}(n)$  was a mediant sum of values of the form  $p_i \log(p_i/(p_i-1))$  ( $\frac{a}{b}$  and  $\frac{c}{d}$  give a mediant of  $\frac{a+c}{b+d}$ ), so that  $(q_k/(q_k-1))^{q_k} < (1/\pi^{-1}(n))^{1/\sigma^{-1}(n)} < (q_1/(q_1-1))^{q_1}$ . Also, if  $\sigma^{-1}(n) \ge 1$ , then  $(1/\pi^{-1}(n))^{1/\sigma^{-1}(n)} < 3.6$ , so one can improve the constant C in  $g(n) \le Ak^{B+C\log\log k}$  to a value less than 3.81. If we did not care about the advantage given by using  $\sigma^{-1}(n)$ , we could use a general bound of  $g(n) < k^{3+3.81\log\log k}$  for k > 2, which can be verified by hand for small values of k and which would be weaker (and thus valid) than the Variation when  $\sigma^{-1}(n) \le 1$ .

Letting  $n=P_k$ , Mertens determines  $\sigma^{-1}(n)$  and  $\pi^{-1}(n)$  with error by  $\sigma^{-1}(n)=\log\log p_k+B+E_1(k)$  and  $\pi^{-1}(n)=e^{-(\gamma+\delta(k))}/\log p_k$  where B is a constant with value near 0.2615 and  $E_1(k)$  and  $\delta(k)$  are error terms in  $O(1/\log p_k)$ . In using Stevens's argument with tighter bounds, we used  $1/2+\log\log p_k$  as an upper bound for  $\sigma^{-1}(n)$ ; calculations show that  $\sigma^{-1}(n)<0.41+\log\log p_k$  for 1/20, and Mertens estimates of the error (or tighter estimates given by Rosser and Schoenfeld) show this holds for 1/20. If we were concerned only with 1/20, we could replace 0.41 by 0.28, closer to the value of 1/21. We could also use 1/22 hog 1/23 as an upper bound for 1/24 by smaller numbers for 1/25 sufficiently large.

Similarly  $\pi^{-1}(n) \log(p_k)$  approaches  $e^{-\gamma}$  which is near 0.5615; computing small examples shows  $\pi^{-1}(n) > 1/3 \log p_k$  for  $1 < p_k < 10^8$ ; theory then gives the weak inequality for all k. We could use  $1/2 \log p_k$  for k > 8 if we needed to improve the multiplicative constant 4; instead we chose to develop the estimate depending on  $\sigma^{-1}(n)$  in the section following Stevens's argument.

The rest of this Appendix contains the verification of the claim that the choice of  $q_l$  for n with  $k > e^{9.5}$  satisfies  $q_l < k^{0.45}$ , and remarks expanding on the proof idea of the Improvement.

We first work with sums of the form  $\sum_{1 \leq j \leq u} 1/p_{m+j}$  which are within  $1/2p_m$  to 1, as they represent an extremal form with respect to the estimate.  $\sum_{4 \leq j \leq 29} 1/p_i < .9$ , so when  $p_m = 5$  we already have  $u > p_m^2$  if we want a sum close enough to 1. As one increases m by 1, one has to remove  $1/p_{m+1}$  and "replace" it by more than  $2p_m$  terms of size smaller than  $1/4p_m^2$ , so we already have  $m/u < 1/m \log m$  for such sums.

As m grows,  $\log p_{m+u}$  tends to  $e \log p_m$  and  $(m+u) \log(m+u)$  approaches  $p_m^e$ , and thus  $u/p_m^e$  approaches 1, yielding the asymptotic (in k) result of  $\log(g(n)) < k^{1/e+\epsilon}$ . Because of oscillations around zero of the quantity  $(\sum_{1 \leq j \leq u} 1/p_j - B - \log \log p_u)$ , a proof of  $q_{l-1} \leq p_m < k^{0.45}$  seems more challenging; we will use results of Rosser and Schoenfeld [9] to show this bound for special sums of the above form for  $m \geq 184$ , then show how computations bring m down to 20, and then show how this implies the general result when  $k > e^{9.5}$ .

We start with getting  $\log p_{m+u}$  in terms close to  $\log u$ .

**Lemma:**  $\log p_{m+u} < \log u + (\log \log u + \frac{m}{u})(1 + \frac{1}{\log u}) + \frac{m}{u(\log u)^2}$  when u > m.

Sketch of Proof: Theorem 6 of [9] yields for k > 5 that  $\log p_k < \log k + \log(\log k + \log\log\log k)$ . We start with the more complicated subterm:  $\log(\log(m+u) + \log\log(m+u)) = \log\log u + LL$ , where  $LL = \log(\log(m+u)/\log u + \log\log(m+u)/\log u)$ .

```
LL = \log(1 + [\log(1 + m/u) + \log\log(m + u)]/\log u)
< [m/u + \log\log(m + u)]/\log u
= [m/u + \log(\log(u) + \log(1 + m/u))]/\log u
= [m/u + \log\log u + \log(1 + \log(1 + m/u)/\log u)]/\log u
< [m/u + \log\log u + m/u\log u]/\log u
= (m/u + \log\log u)/\log u + m/u(\log u)^{2}
```

Now we use Theorem 6 to get

$$\log p_{m+u} < \log(m+u) + \log(\log(m+u) + \log\log(m+u))$$

$$< \log u + m/u + \log\log u + LL$$

$$< \log u + (m/u + \log\log u)(1 + 1/\log u) + m/u(\log u)^{2}.$$

#### End of Sketch.

Next, we want to get a good lower bound on  $\log(\log p_{m+u}/\log p_m)$  when we have our special sum close enough to 1.

**Lemma:** Suppose  $\sum_{1 \le j \le u} 1/p_{m+j} > 1 - 1/2p_m$  and  $p_{m+u} > 286$ . Then  $\log(\log p_{m+u}/\log p_m) > 1 - 1/2p_m - .00148 - 1/2(\log p_{m+u})^2$ .

Sketch of Proof: From [9] Theorem 5 we derive (where  $p \le x$  means the primes greater than 1 and at most x)

$$|\sum_{p \le x} 1/p - \log \log x - B| \le 1/2(\log x)^2 \text{ for all } x \ge 286$$

and from [9] Theorem 20 we have

$$\sum_{p \le x} 1/p - \log \log x - B > 0 \text{ for } 2 \le x \le 10^8.$$

Combining the results over the two ranges gives

$$-\log\log p_m - B + 1/2(\log(10^8))^2 > = -\sum_{p \le p_m} 1/p \text{ for all } m > 0$$

leading to our estimate:

$$\sum_{p_m 
$$-\log \log p_m - B + 1/(2(\log 10^8)^2)$$

$$\le \log(\log p_{m+u}/\log p_m) + \frac{1}{2(\log p_{m+u})^2} + .00148$$$$

Subtracting the last two terms from  $1-1/2p_m$  leads to the conclusion. **End** of Sketch.

Now let us choose  $p_m > e^7 > 1096$  and so  $m \ge 184$  and  $1/\log u < 1/15$ . Then  $\log(\log p_{m+u}/\log p_m) > 1 - 1/2192 - .00148 - 1/2(15)^2 > 0.99$  and  $e^{0.99} > 2.691$ , thus  $\log p_{m+u} > 2.691 \log p_m$ .

Toward a contradiction, assume that  $x = \frac{20 \log p_m}{9} > \log u$ . Then the Lemma concerning  $\log p_{m+u}$  gives

1.21095
$$x = 2.691 \log p_m < \log p_{m+u}$$
  
 $< x + (\log x + m/u)(1 + 1/\log u) + m/u(\log u)^2$   
 $< x + (\log x + 1/1000)(1 + 1/15) + 1/1000$   
 $< x + (16 \log x)/15 + 1/100$ 

However  $.21095x > (16 \log x)/15 + .01$  for x > 14, which means for  $\log p_m > 7$ , and we have a contradiction. Thus  $u \ge p_m^{20/9}$ .

For primes  $p_m=71$  through  $p_m=857$  (m=20 to m=148), we verified through computation that if  $\sum_{1\leq j}^u 1/p_{m+i}>1-1/2p_m$ , then  $\log u/\log p_m>20/9$ . In particular for u as low as  $13250< e^{9.5}$  and n a ratio of certain primorials the Improvement holds. Also,  $\sum_{1\leq j}^{5761308} 1/p_{148+j}>1-1/2p_{148}$ , thus  $\sum_{1\leq j}^u 1/p_{m+j}>1-1/2p_m$  implies  $u>p_m^{20/9}$  for  $p_m$  running up to  $1102<5761308^{9/20}$ . So the implication holds for all  $p_m\geq 71$ .

Now assume n with  $k \ge e^{9.5}$  and otherwise arbitrary. Find l smallest so that  $\sum_{l \le i}^k 1/q_i < 1 + 1/2q_l$ . Let  $p_m = q_{l-1}$ . Then  $1 - 1/2q_{l-1} \le \sum_{l \le i}^k 1/q_i \le \sum_{1 \le j}^{k-l+1} 1/p_{m+j}$ . Either  $p_m \ge 71$  and so  $(k-l+1)^{0.45} \ge u^{0.45} \ge p_m$ , or else  $q_{l-1} < 71 < (e^{9.5})^{9/20} \le k^{0.45}$ .

# 9 Addendum

Shortly after version 1 of this article was posted, we found an upper bound based on estimates of Euler's totient which does better than the Improvement.

Aaron Meyerowitz asked about these estimates in question 88777 on MathOverflow. The key observation is that the number of totatives to n in the interval [0,x) differs from  $x\phi(n)$  by a periodic function  $E_n(x)$  which has a maximum value at most  $2^{k-1}$ . Using this, we can generalize the Variation using a similar argument to show

$$g(n) \le \frac{(k-m+1)(2^m + \pi^{-1}(d))}{\pi^{-1}(d)(1-\sigma^{-1}(n/d))},$$

where one chooses m < k so that  $d = \prod_{1 \le i \le m} q_i$  divides n and also satisfies  $t = 1 - \sigma^{-1}(n/d) > 0$ , and finally so that the right hand side above leads to an optimal bound. (Hint: the Proof sketch of

the Variation introduces an error term of size at most 2 for counting multiples of  $q_i$  which are not multiples of  $q_1$ ; now use an error of  $2^m + \pi^{-1}(d)$  for an interval [y, x] of length L when counting multiples of  $q_i$  that are coprime to d.)

When such an optimal m > 0 is found, one can easily show that  $t > 1/(3 + q_{m+1})$ , and the analysis from the Appendix can be adapted to show  $m^e < k$ . As  $1/\pi^{-1}(d)$  is bounded by  $2 \log q_m$ , this will beat our bound  $k^{K\sigma^{-1}(n)}$  when  $2^m$  is substantially smaller than  $k^{K\sigma^{-1}(n)-3/2}$ . Computing both bounds for  $n = P_k$  for  $k < 10^7$  show this new bound to be superior: we expect to show that it holds when  $k < 10^{10}$  and  $\sigma^{-1}(n) > 1$ , which would imply that K above can be taken near 3.6 for all k > 2.

This bound can be used to show Kanold's bound  $2^{\sqrt{k}}$  actually holds for  $k > e^6$ . There are also improvements to be made on the error term  $(2^m + \pi^{-1}(d))$ ; with such improvements we expect to show in a followup article a subquadratic in k upper bound for  $k < 10^{10}$ .

We also found a statement of the Observation in a 1975 work of Kanold's. (We thank Prof. Dr. Heiko Harborth for making this part of the literature available to us.) We are still looking for a published proof of the Observation as well as an appearance in the literature of the Proposition that  $g(n) \leq \lceil k/(1-\sigma^{-1}(n)) \rceil$  for those n with  $\sigma^{-1}(n) < 1$ . We still believe that the explicit upper bound of  $k^{3+3.81\sigma^{-1}(n)}$  has not appeared in the literature.

We have planned a series of forthcoming articles, tentatively titled 'Adventures in finding bounds on Jacobsthal's function.' In addition to fleshing out some of the questions asked in an earlier section, we will consider the computational complexity of g(n) and approximations to g(n), various lower bounds coming from elementary (and not so elementary) considerations, applications using both conjectured and actual bounds, and generalizations in geometric and algebraic realms.

#### References

- [1] Erik Westzynthius, Über die Verteilung der Zahlen, die zu den n ersten Primzahlen teilerfremd sind. Societas Scientiarum Fennica Commentationes Physico-Mathematicae V. 25, Helsingfors 1931, pp. 1-37.
- [2] Ernst Jacobsthal, Über Sequenzen ganzer Zahlen, von denen keine zu n teilerfremd ist. I. Det Kongelige Norske Videnskabers Selskabs

- Forhandlinger Bd 33 1960, Nr. 24, Trondheim I Kommisjon Hos F. Bruns Bokhandel 1961, pp. 117-124.
- [3] Hans-Joachim Kanold, Über eine zahlentheoretische Funktion von Jacobsthal. *Mathematische Annalen* 170 1967, pp.314-326.
- [4] Harlan Stevens, On Jacobsthal's g(n)-function. Mathematische Annalen 226 1977, pp. 95-97.
- [5] Thomas R. Hagedorn, Computation of Jacobsthal's function h(n) for n < 50. Mathematics of Computation 78 2009, pp. 1073-1087
- [6] MathOverflow question Erik Westzynthius's cool upper bound argument: update? http://mathoverflow.net/questions/37679/
- [7] Gerhard Paseman, The Waltraud and Richard R. Paseman Theorem, private manuscript, March 2011
- [8] L. Hajdu and N. Saradha, Disproof of a conjecture of Jacobsthal. Mathematics of Computation V 81 n. 280 October 2012, pp. 2461-2471.
- [9] J. Barkley Rosser, Lowell Schoenfeld, Approximate formulas for some functions of prime numbers. *Illinois Journal of Mathematics* 6 1962, pp 64-94.