RESEARCH STATEMENT – NATHAN McNEW

1. Summary of Thesis Work

My research is focused primarily on problems in analytic number theory, especially those problems where analytic techniques can be used to answer number theoretic questions of a combinatorial or probabilistic nature. In my thesis I improve upon existing results regarding three general problems: Counting a subset of integers closely related to the Carmichael and Lehmer numbers; the Ramsey-theoretic problem to study the densities of large subsets of the integers which avoid geometric progressions; and the study of the distribution of the largest prime factor of the integers up to x, determining in particular which prime occurs most frequently.

1.1. **Radically weakening the Carmichael and Lehmer conditions.** Let $\varphi(n)$ denote the Euler totient function of n and $\lambda(n)$ the size of the largest cyclic subgroup of $(\mathbb{Z}/n\mathbb{Z})^\times$. Lehmer [15] asked whether there exist composite integers n such that $\varphi(n)|n-1$. No "Lehmer numbers," integers satisfying this condition, are known. Luca and Pomerance [16] showed the count of Lehmer numbers up to x is $O_\epsilon\left(\frac{x^{1/2}}{(\log x)^{1/2+\epsilon}}\right)$. Carmichael numbers are composite numbers, n, which satisfy Fermat's little theorem for every base a; that is, $a^n \equiv a \pmod n$. Carmichael [7] showed these integers satisfy a weakening of the Lehmer condition, namely a composite integer n is Carmichael if $\lambda(n)|n-1$. (Note $\lambda(n)|\varphi(n)$ for all n.) Unlike Lehmer numbers, it is known [1] there are infinitely many Carmichael numbers. Pomerance [29], [28] gives an upper bound for the number, $C(x)$, of Carmichael numbers up to x as well as a heuristic that this bound is the true size of $C(x)$, namely

$$C(x) \leq x^{1-\{1+o(1)\}\log\log\log x/\log\log x}. \tag{1.1}$$

Grau and Oller-Marcén [13] suggest another possible weakening of the Lehmer property: looking at the set $\mathbb{L}_k$ of composite n such that $\varphi(n)|(n-1)^k$ as well as the union, $\mathbb{L}$, of all the $\mathbb{L}_k$; that is the set of n where every prime dividing $\varphi(n)$ also divides $n-1$. This last set is a weakening of both the Lehmer and Carmichael conditions, since every prime dividing $\varphi(n)$ also divides $\lambda(n)$.

In [19] I resolve several conjectures of Grau and Oller-Marcén by showing that the count, $L(x)$, of the count of integers in $\mathbb{L}$ up to x satisfies the same upper bound (1.1) as $C(x)$. I also give upper bounds for the count of integers in $\mathbb{L}_k$ up to x for each $k > 1$ and the count of those integers in $\mathbb{L}$ which are the product of a fixed number of primes.

In that paper I conjecture, based on strong numerical evidence, that there are infinitely many members of $\mathbb{L}$ which are not Carmichael. Utilizing the recent progress by Maynard [17] and Tao [26] on the prime k-tuples conjecture, Tom Wright and I, in very recent work [21], show this is in fact the case. Moreover, we show that $L(x) \gg x^{\frac{1}{2}-o(1)}$ (note the best known lower bound for the Carmichael numbers is $C(x) \gg x^{\frac{1}{3}}$), that there are infinitely many members of $\mathbb{L}$ with any fixed number (at least 2) of prime factors, and that $\mathbb{L}_k$ contains infinitely many integers for each $k \geq 3$.

1.2. **Avoiding geometric progressions in the integers.** A well known problem in combinatorial number theory is to study sets of integers which do not contain an arithmetic progression (**AP**). For $A \subset \mathbb{N}$, denote its density by $d(A) = \lim_{N\to\infty} \frac{A\cap[1,N]}{N}$ (if the limit exists), and upper density by $\bar{d}(A) = \limsup_{N\to\infty} \frac{A\cap[1,N]}{N}$. Roth famously proved that if

$\bar{d}(A) > 0$ then A contains a 3-term AP. His argument shows further [32] that any subset of $\mathbb{Z}/n\mathbb{Z}$ free of 3-term APs has size $O\left(\frac{n}{\log\log n}\right)$. His result has been improved many times; Sanders [34] improved it to $O\left(\frac{N(\log\log n)^5}{\log n}\right)$. Szemerédi [36] and Furstenberg [12] independently generalize Roth's result to arbitrarily long APs.

In a completely analogous manner, one can consider a geometric progression (**GP**) of integers of the form (a, ak, ak^2) for $k \in \mathbb{Q}$ and seek sets which are free of such progressions. It is surprising just how different the results are in this case. This problem was first considered by Rankin [30] who showed that there exists a set, G_3^*, free of geometric progressions with asymptotic density $d(G_3^*) = \frac{1}{\zeta(2)}\prod_{i>0}\frac{\zeta(3^i)}{\zeta(2\cdot 3^i)} \approx 0.71974$. (Here ζ is the Riemann zeta function.) We will use the notation $\overline{\alpha} = \sup\{\bar{d}(A) : A \subset \mathbb{N} \text{ is GP-free}\}$ and $\alpha = \sup\{d(A) : A \subset \mathbb{N} \text{ is GP-free and } d(A) \text{ exists}\}$ for sets free of progressions with rational ratio and the constants β and $\overline{\beta}$ defined analogously for sets which avoid progressions with integral ratio. Thus $d(G_3^*) \leq \alpha \leq \overline{\alpha}, \beta \leq \overline{\beta}$.

Many authors have studied upper bounds for $\overline{\alpha}$. Riddell [31] showed that $\overline{\alpha} \leq 6/7$. This bound was reproved by Beiglböck, Bergelson, Hindman, and Strauss [4], who were unaware of Riddell's result. Their result was an improvement of a bound by Brown and Gordon [6] who, also unaware of Riddell's result, had shown $\overline{\alpha} < 0.86889$. Nathanson and O'Bryant [24] combined these methods to show $\overline{\alpha} < 0.84948$. Beiglböck, et al. [4] appear to have been the first to consider $\overline{\beta}$, and show that $\overline{\beta} \geq 0.75$.

In [20] I improve on many of the results above. I demonstrate an algorithm to effectively compute the values of both $\overline{\alpha}$ and $\overline{\beta}$ and use this to significantly improve the best known bounds for each. In particular, I show $0.730027 < \overline{\alpha} < 0.772059$ and $0.815509 < \overline{\beta} < 0.819222$. The lower bound for $\overline{\beta}$ comes from a different construction which utilizes the fact that a geometric progression with integral ratio cannot have more than one term in a short interval. This construction has since been pursued further by Nathanson and O'Bryant [25] to study sets of real numbers avoiding progressions with integral ratio. Ford improves upon my construction by carefully choosing integers in one of the intervals to be coprime to small primes, thus obtaining the bound $0.818410 < \overline{\beta}$.

Thus far Rankin's set, G_3^*, obtained by greedily including integers which avoid progressions with rational ratio, has the greatest asymptotic density of sets avoiding rational-ratio progressions. Interestingly, if one restricts to integer ratios and constructs the set greedily avoiding such progressions, the resulting set is the same as for rational ratios. Prior to my work this greedy set also had the highest known density among sets avoiding integral ratios, however I construct a set which does so with a slightly greater density, showing that $0.72195 < \beta$. Towards an upper bound for α or β, I use the logarithmic density to show that a set with asymptotic density cannot do better than the greedy set while avoiding ratios which are a power of two. This greedy set has an asymptotic density approximately 0.84539, strictly less than the optimal upper density, about 0.84638.

In the spirit of Roth's theorem for subsets of $\mathbb{Z}/n\mathbb{Z}$ avoiding arithmetic progressions, I consider the problem of finding subsets of $\mathbb{Z}/n\mathbb{Z}$ which avoid geometric progressions. In contrast to the results over $\mathbb{N}$, I show that it is not possible to take a positive proportion of the residues modulo n in this case. In fact the largest such subset has size $O\left(\frac{n(\log\log n)^5}{(\log n)^{1/2}}\right)$.

1.3. **Finding the most popular values of the largest-prime-divisor function.** Let $P(n)$ denote the largest prime factor of n. Several authors have investigated the distribution

of $P(n)$ for $n \in [2,x]$. Erdős and Alladi [3] showed that the mean is given by $\frac{1}{x}\sum_{n=2}^{x} P(n) = (1+o(1))\frac{\pi^2 x}{12\log x}$. De Koninck and Ivić [10] give an expression with lower order terms, (Naslund [22] worked out explicit values of the d_m)

$$\frac{1}{x}\sum_{n=2}^{x} P(n) = \frac{\pi^2 x}{12\log x} + \frac{d_2 x}{\log^2 x} + \cdots + \frac{d_m x}{\log^m x} + O\left(\frac{x}{\log^{m+1} x}\right).$$

The median value of $P(n)$, for $n \in [2,x]$, given as $x^{1/\sqrt{e}+o(1)}$, first appeared in a paper of Selfridge and Wunderlich [35] in 1974, though this fact had been known since at least the early 20th century. Naslund [23] showed that this median is given more accurately by $e^{\frac{\gamma-1}{\sqrt{e}}} x^{\frac{1}{\sqrt{e}}}\left(1+O\left(\frac{1}{\log x}\right)\right)$. The fact that the median grows so much slower than the mean is somewhat indicative of just how strongly right skewed this distribution is. In fact, letting $f_x(p) = \#\{n \le x : P(n) = p\}$, De Koninck showed [9] that for fixed x the maximum value of $f_x(p)$ is attained by a prime p satisfying

$$p = e^{\sqrt{\frac{1}{2}\log x(\log\log x + \log\log\log x)} + O\left(\sqrt{\frac{\log x}{\log\log x}}\right)}.$$

Furthermore, he and Sweeney show [11] that $f_x(p)$ is increasing on the primes in the interval $[2,\sqrt{\log x}]$ and decreasing for $[\sqrt{x},x]$. I am interested in better understanding the behavior in between these two intervals, and in particular the location of the maximum value, as well as the set of primes which achieve this maximum for some value of x.

In ongoing work [18] I show that $f_x(p)$ is maximized by a prime p satisfying

$$p = e^{\sqrt{\upsilon(x)\log x} + O\left((\log\log x)^{1/4}\right)} \tag{1.2}$$

where $\upsilon(x)$ is the solution to $e^{\upsilon(x)} = \sqrt{\upsilon(x)\log x - \upsilon(x)^2}$. Standard techniques of asymptotic approximation give that $\upsilon(x) = \frac{1}{2}\log\log x + \frac{1}{2}\log\log\log x - \frac{1}{2}\log 2 + o(1)$. As the location of this maximum value tends, slowly, to infinity, one might expect that every prime is the maximum value of $f_x(p)$ for some x. This turns out not to be the case, for example 11 is not at the peak of this distribution for any x. Define a **popular prime** to be a prime which achieves the maximum value. In particular, a prime p is popular if there exists an N for which p is the mode of $P(n)$ for $n \in [2,N]$.

Not only are there primes, like 11, which are not popular, but in fact the popular primes have relative density zero in the primes. I use analytic techniques to show that if p_n and p_{n+k}, the primes indexed by n and $n+k$ respectively, are both popular then $\frac{1}{k}(p_{n+k} - p_n) = \log p_{n+k} + O(\log\log p_{n+k})$. In other words, the difference between p_n and p_{n+k} must be remarkably close to the average spacing. When $k < (\log p_n)^{2/3}$ is small, finding primes this close to the average spacing is very unusual and so, using sieve techniques, I show that the number of popular primes up to x is $O\left(\frac{x}{(\log x)^{4/3+o(1)}}\right)$. On the other hand, due to the error term in (1.2) this count is at least $C\frac{\log x}{(\log\log x)^{1/4}}$ for some constant C. There clearly remains a huge gap between these bounds!

2. Future directions

The work described above can be extended in a variety of ways, and there remains several important unsolved problems. Regarding the first problem, generalizing the Carmichael and Lehmer conditions, I conjecture in [19] not only that the difference $L(x) - C(x)$ of the count

of the number of Carmichael numbers from the integers in $\mathbb{L}$ tends to infinity (we answer this in the affirmative in [21]) but that the ratio $\frac{L(x)}{C(x)}$ does as well. This question will likely be much harder, however it would be interesting to obtain better lower bounds for $L(x)$ and to better understand those members of $\mathbb{L}$ which do not arise from the prime k-tuples construction, which should account for most of $\mathbb{L}$.

It may be possible to further exploit the recent results on primes in tuples to study the sets $\mathbb{L}_k$. It remains open whether $\mathbb{L}_2$ is infinite and bounds for the sizes of the $\mathbb{L}_k$ are far from optimal. Techniques used by other authors to give upper bounds for the Lehmer numbers might be useful here. It would also be nice to understand how Carmichael numbers are distributed in the sets $\mathbb{L}_k$.

Regarding sets avoiding geometric progressions, I list several open problems in [20] which would be very interesting to pursue further. While the constants $\overline{\alpha}$ and $\overline{\beta}$ are now comparatively well understood, α and β are not. The best upper bound for each remains the bounds for $\overline{\alpha}$ and $\overline{\beta}$, though presumably the true values are significantly smaller. It would be interesting to see if there is a way the arguments I develop using logarithmic density to study sets avoiding ratios that are powers of two could be extended to prove that α is strictly less than $\overline{\alpha}$. Likewise for β and $\overline{\beta}$. While there is reason to believe Rankin's set gives the true value of α, I have shown this is not the case for β by giving an explicit construction of a set with higher density. This example was found computationally, however, so there remains much to be understood as to how one can tweek Rankin's set slightly to create such examples, and what the limit of such constructions might be.

Another natural question is to consider geometric progressions of length greater than 3. Many authors already study progressions of arbitrary length, however I chose for the sake of clarity to focus on the case of 3-geometric progressions in my paper. While the algorithms I describe to compute $\overline{\alpha}$ and $\overline{\beta}$ should apply to progressions of arbitrary length, other results will be more complicated. For example, as Riddell [31] discusses, the structure of the greedy set is far more complicated for composite-length-progressions: the construction given by Rankin [30] is not optimal in that case.

In [25] Nathanson and O'Bryant use the construction I give for large sets of integers avoiding integral ratios using intervals to study subsets of the real numbers which avoid geometric progressions. They consider progressions of length greater than 3 and give a table of the optimal intervals to take for progressions of length up to 9 and denominators up to 10,000. However "Ford's trick" of taking the integers in certain intervals to be coprime to small primes has no analogue over the reals, and it remains to be seen if it is useful when avoiding longer integral ratio progressions in $\mathbb{Z}$.

One can also consider geometric progressions in more general rings. In the summer of 2014 I worked with a team of undergraduate students at the Williams College REU [5] to generalize my work to the ring of integers over a number field. We characterized the greedy sets of integers (or ideals in a non-UFD) for quadratic number fields and give bounds for the upper density of a set of algebraic integers (or ideals) avoiding 3 term geometric progressions which depend on the splitting behavior of the small primes. This work relies critically on the fact that we have unique factorization in the ring of integers (or ideals) in the number field. It is not yet clear, for example, what the greedy set of algebraic integers (not ideals) avoiding geometric progressions in an imaginary quadratic number field with class number greater than one would look like. One could likewise look at geometric progressions in other

rings where useful properties of the integers no longer hold, for example in $SL_2(\mathbb{Z})$ or the Hurwitz order of integral quaternions where commutativity fails.

This sort of problem is well suited for undergraduate research as it can be easily motivated and the combinatorial techniques involved can be picked up relatively easily by advanced undergraduate students while introducing them to concepts, like number fields, Euler products, and unique factorization, they are unlikely to encounter otherwise. I plan to continue involving undergraduates in as many aspects of this research as possible.

Finally, the work on the popular primes is rich with potential for further research. The present upper and lower bounds for their count clearly leave much room for improvement. These bounds appear to be the best one can obtain using the work of Hildebrand [14] and Alladi [2] to estimate the count, $\Psi(x,y)$, of y-smooth integers up to x. Saias [33] gives a much better approximation of $\Psi(x,y)$ in terms of a function, $\Lambda(x,y)$, first introduced by de Bruijn. This function is notoriously difficult to work with, however if its behavior could be better understood near the critical value of y for a given x, it could lead to a much better understanding of the popular primes.

One can also consider generalizations of the popular primes. De Koninck considered the most popular value of the k-th largest prime divisor, and showed it is always 3 for all $k \geq 2$ and sufficiently large x. On the other hand, one could consider the second most popular values of $P(n)$ and the set of primes, the "runner up primes", which achieve this rank without being popular. This readily generalizes to any fixed rank. Assign a "popularity index" to each prime corresponding to the highest rank it achieves in the count of the largest prime divisors on $[2,x]$ for any value of x. Thus, the popular primes would be precisely those primes with popularity index one. It would be highly interesting to see if the primes with bounded rank are sparsely distributed within the primes, or if there is a radical change in behavior when going from $k=1$ to $k \geq 2$ as in De Koninck's problem.

Determining which primes are in fact popular is a computationally difficult problem. As far as we can compute so far they contain all of the convex primes, which are the vertices of the convex hull of the points $(n,p_n) \in \mathbb{R}^2$. Pomerance [27] showed the count of the convex primes up to x is at least $\exp\left(c(\log x)^{3/5}\right)$ for some $c>0$. He used this to show that the set of primes which satisfy the inequality $p_{n-i}+p_{n+i} > 2p_n$ for all $i<n$ (of which the convex primes are a subset) is also infinite. Thus far every popular prime except 773 also has this property. It would be interesting to better understand the connection between these sets, especially since the convex primes are far easier to count.

A step in many factoring algorithms is to choose y-smooth integers up to x until the product of a subset of them is a square. The expected number of trials to pick a y-smooth integer is $\frac{x}{\Psi(x,y)}$ and having $\pi(y)+1$ such integers forces a square dependence, so the optimal value of y minimizes $\frac{x\pi(y)}{\Psi(x,y)} \approx \frac{xy}{\Psi(x,y)}$ or equivalently, maximizes $\frac{\Psi(x,y)}{y}$.

The analysis of this optimal smoothness bound is similar to that of the popular primes, and in fact it is known [8] that the optimal smoothness bound satisfies an expression very similar to equation (1.2). As such, the study of the popular primes may have applications to the optimization of such factoring algorithms. In particular it could be useful to compare the set of popular primes (which maximize $\Psi\left(x/p,p\right)$ for some x) to the set of those "fast primes," p, which maximize $\Psi(x,p)/p$ for some value of x. In [8] , the authors just fail to get an asymptotic, and so far I don't have one either for popular primes (see (1.2)). Can both problems be solved with the same technique?

References

[1] W. R. Alford, A. Granville, and C. Pomerance. There are infinitely many Carmichael numbers. *Ann. of Math. (2)*, 139(3):703–722, 1994.

[2] K. Alladi. The Turán-Kubilius inequality for integers without large prime factors. *J. Reine Angew. Math.*, 335:180–196, 1982.

[3] K. Alladi and P. Erdős. On an additive arithmetic function. *Pacific J. Math.*, 71(2):275–294, 1977.

[4] M. Beiglböck, V. Bergelson, N. Hindman, and D. Strauss. Multiplicative structures in additively large sets. *J. Combin. Theory Ser. A*, 113(7):1219–1242, 2006.

[5] A. Best, K. Huan, N. McNew, S. Miller, J. Powell, K. Tor, and M. Weinstein. Ramsey theory in quadratic number fields. *In preparation.*

[6] B. E. Brown and D. M. Gordon. On sequences without geometric progressions. *Math. Comp.*, 65(216):1749–1754, 1996.

[7] R. D. Carmichael. Note on a new number theory function. *Bull. Amer. Math. Soc.*, 16(5):232–238, 1910.

[8] E. Croot, A. Granville, R. Pemantle, and P. Tetali. On sharp transitions in making squares. *Ann. of Math. (2)*, 175(3):1507–1550, 2012.

[9] J.M. De Koninck. On the largest prime divisors of an integer. In *Extreme Value Theory and Applications*, pages 447–462. Springer, 1994.

[10] J.M. De Koninck and A. Ivić. The distribution of the average prime divisor of an integer. *Arch. Math. (Basel)*, 43(1):37–43, 1984.

[11] J.M. De Koninck and J. Sweeney. On the unimodal character of the frequency function of the largest prime factor. *Colloq. Math.*, 88(2):159–174, 2001.

[12] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.*, 31:204–256, 1977.

[13] J. M. Grau and A. M. Oller-Marcén. On k-Lehmer numbers. *Integers*, 12(A37), 2012.

[14] A. Hildebrand. On the number of positive integers $\leq x$ and free of prime factors $> y$. *J. Number Theory*, 22(3):289–307, 1986.

[15] D. H. Lehmer. On Euler's totient function. *Bull. Amer. Math. Soc.*, 38(10):745–751, 1932.

[16] F. Luca. On composite integers n for which $\varphi(n)|n-1$. *Boletin de la Sociedad Matemática Mexicana*, 17:13–21, 2011.

[17] J. Maynard. Small gaps between primes. *Ann. of Math. (2)*, to appear.

[18] N. McNew. Popular values of the largest prime divisor function. *In preparation.*

[19] N. McNew. Radically weakening the Lehmer and Carmichael conditions. *Int. J. Number Theory*, 9(5):1215–1224, 2013.

[20] N. McNew. Sets of integers which contain no three terms in geometric progression. *Math. Comp.*, To Appear.

[21] N. McNew and T. Wright. There are infinitely many k-Lehmer numbers which are not Carmichael. *In preparation.*

[22] E. Naslund. The average largest prime factor. *Integers*, 13:Paper No. A81, 5, 2013.

[23] E. Naslund. The median largest prime factor. *J. Number Theory*, 141:109–118, 2014.

[24] M. B. Nathanson and K. O'Bryant. On sequences without geometric progressions. *Integers*, 13:Paper No. A73, 5, 2013.

[25] M. B. Nathanson and K. O'Bryant. A problem of Rankin on sets without geometric progressions. *arXiv preprint arXiv:1408.2880*, 2014.

[26] D. H. J. Polymath. Variants of the selberg sieve, and bounded intervals containing many primes. *Res. Math. Sci*, to appear.

[27] C. Pomerance. The prime number graph. *Math. Comp.*, 33(145):399–408, 1979.

[28] C. Pomerance. On the distribution of pseudoprimes. *Math. Comp.*, 37(156):587–593, 1981.

[29] C. Pomerance, J. L. Selfridge, and S. S. Wagstaff Jr. The pseudoprimes to $25 \cdot 10^9$. *Math. Comp.*, 35(151):1003–1026, 1980.

[30] R. A. Rankin. Sets of integers containing not more than a given number of terms in arithmetical progression. *Proc. Roy. Soc. Edinburgh Sect. A*, 65:332–344 (1960/61), 1960/1961.

[31] J. Riddell. Sets of integers containing no n terms in geometric progression. *Glasgow Math. J.*, 10:137–146, 1969.

[32] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.
[33] É. Saias. Sur le nombre des entiers sans grand facteur premier. *J. Number Theory*, 32(1):78–99, 1989.
[34] T. Sanders. On Roth's theorem on progressions. *Ann. of Math. (2)*, 174(1):619–636, 2011.
[35] J. L. Selfridge and M. C. Wunderlich. An efficient algorithm for testing large numbers for primality. In *Proceedings of the Fourth Manitoba Conference on Numerical Mathematics (Winnipeg, Man., 1974)*, pages 109–120. Congr. Numer., No. XII. Utilitas Math., Winnipeg, Man., 1975.
[36] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975. Collection of articles in memory of Juriĭ Vladimirovič Linnik.