Honors 1: Undergraduate Math Lab*

Peter Sarnak†, Steven Miller‡, Alex Barnett§

October 23rd, 2002

Courant Institute of Mathematical Sciences
New York University
New York, NY

Abstract

Introduction to the Hardy-Littlewood Circle Method. Lecture and notes by Steven Miller.

1 Problems where the Circle Method is Useful

For each N, let A_N be a set of non-negative integers such that

1. $A_N \subset A_{N+1}$,
2. $|A_N| \to \infty$ as $N \to \infty$.

Let $A = \lim_{N\to\infty} A_N$.

Question 1.1. *Let s be a fixed positive integer. What can one say about $a_1 + \cdots + a_s$? Ie, what numbers n are representable as a sum of s summands from A?*

We consider three problems; we will mention later why we are considering sets A_N.

*Homepage: `http://www.math.nyu.edu/~millerj/`

†E-mail: `sarnak@math.princeton.edu`
‡E-mail: `millerj@cims.nyu.edu` or `sjmiller@math.princeton.edu`
§E-mail: `barnett@nmr.mgh.harvard.edu`

1.1 Waring's Problem

Let A be the set of k^{th} powers of non-negative numbers, and let

$$A_N \;=\; \{0^k, 1^k, \ldots, N^k\}. \tag{1}$$

Question 1.2. *Fix a positive integer k. For what positive integers s can every integer be written as a sum of s numbers, each number a k^{th} power?*

Thus, in this case, we are trying to solve

$$n \;=\; a_1^k + \cdots + a_N^k. \tag{2}$$

1.2 Goldbach's Problem

Let A be the set of all prime numbers, and let A_N be the set of all primes at most N.

Question 1.3. *Can every even number be written as the sum of two primes?*

In this example, we are trying to solve

$$2n \;=\; a_1 + a_2, \tag{3}$$

or, in more suggestive notation,

$$2n \;=\; p_1 + p_2. \tag{4}$$

1.3 Sum of Three Primes

Again, let A be the set of all primes, and A_N all primes up to N.

Question 1.4. *Can every odd number be written as the sum of three primes?*

Again, we are studying

$$2n + 1 \;=\; p_1 + p_2 + p_3. \tag{5}$$

2 Idea of the Circle Method

2.1 Introduction

Definition 2.1 ($e(z)$)**.** *We define* $e(z) = e^{2\pi iz}$.

Exercise 2.2. *Let* $m, n \in \mathbb{Z}$. *Prove*

$$\int_0^1 e(nx)e(-mx)dx = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

Let A, A_N be as in any of the three problems above. Consider

$$f_N(x) \;=\; \sum_{a \in A_N} e(ax). \tag{7}$$

We investigate $\Big(f_N(x)\Big)^s$:

$$\begin{aligned} \Big(f_N(x)\Big)^s &= \prod_{j=1}^{s} \sum_{a_j \in A_N} e(a_j x) \\ &= \sum_m r_N(m) e(mx). \end{aligned} \tag{8}$$

The last result follows by collecting terms. When you multiply two exponentials, you add the exponents.

Thus, when we multiply the s products, how can we get a product which gives $e(mx)$?

We have s products, say $e(a_1x)$ through $e(a_Nx)$. Thus,

$$e(a_1x)\cdots e(a_Nx) \;=\; e\Big((a_1 + \cdots + a_N)x\Big) \;=\; e(mx). \tag{9}$$

Thus, the coefficient $r_N(m)$ in $\Big(f_N(x)\Big)^s$ is the number of ways of writing

$$m \;=\; a_1 + \cdots + a_N, \tag{10}$$

with each $a_j \in A_N$.

As the elements of A_N are non-negative, if N is sufficiently large $r_N(m)$ is equal to the number of ways of writing m as the sum of s elements of A.

The problem is, if m is larger than the largest term in A_N, then there may be other ways to write m as a sum of s elements of A.

Lemma 2.3.

$$r_N(m) \;=\; \int_0^1 \Big(f_N(x)\Big)^s e(-mx)dx. \tag{11}$$

Proof: direct calculation.

Note that, just because we have a closed form expression for $r_N(m)$, this does not mean we can actually *evaluate* the above integral. Recall, for example, the inclusion - exclusion formula for the number of primes at most N. This is an exact formula, but very hard to evaluate.

2.2 Useful Number Theory Results

We will use the following statements freely:

Theorem 2.4 (Prime Number Theorem). *Let $\pi(x)$ denote the number of primes at most x. Then*

$$\pi(x) \;=\; \sum_{p\le x} 1 \;=\; \frac{x}{\log x} \;+\; \textit{smaller}. \tag{12}$$

Upon applying Partial Summation, we may rewrite the above as

$$\sum_{p\le x} \log p \;=\; x \;+\; \textit{smaller}. \tag{13}$$

Theorem 2.5 (Siegel-Walfisz). *Let $C, B > 0$, and let a and q be relatively prime. Then*

$$\sum_{\substack{p\le x\\ p\equiv a(q)}} \log p \;=\; \frac{x}{\phi(q)} + O\Big(\frac{x}{\log^C x}\Big) \tag{14}$$

for $q \le \log^B x$, and the constant above does not depend on x, q or a (ie, it only depends on C and B).

For completeness, we include a review of partial summation as an appendix to these notes.

2.3 Average Sizes of $\Big(f_N(x)\Big)^s$

Henceforth we will consider $f_N(x)$ arising from the three prime case. Thus, $s = 3$.

For analytic reasons, it is more convenient to instead analyze the function

$$F_N(x) \ = \ \sum_{p \le N} \log p \cdot e(px). \tag{15}$$

Working analogously as before, we are led to

$$R_N(m) \ = \ \int_0^1 \Big(F_N(x)\Big)^3 e(-mx) dx. \tag{16}$$

By partial summation, it is very easy to go from $R_N(m)$ to $r_N(m)$.

Exercise 2.6. *Prove the trivial bound for $|F_N(x)|$ is N. Take absolute values and use the Prime Number Theorem.*

We can, however, show that the average square of $F_N(x)$ is significantly smaller:

Lemma 2.7. *The average value of $|F_N(x)|^2$ is $N \log N$.*

Proof: The following trivial observation will be extremely useful in our arguments. Let $g(x)$ be a complex-valued function, and let $\overline{g}(x)$ be its complex conjugate. Then $|g(x)|^2 = g(x)\overline{g}(x)$.

In our case, as $\overline{F}_N(x) = F_N(-x)$ we have

$$\begin{aligned} \int_0^1 |F_N(x)|^2 \ &= \ \int_0^1 F_N(x) F_N(-x) dx \\ &= \ \int_0^1 \sum_{p \le N} \log p \cdot e(px) \sum_{q \le N} \log q \cdot e(-qx) dx \\ &= \ \sum_{p \le N} \sum_{q \le N} \log p \log q \int_0^1 e\Big((p-q)x\Big) dx \\ &= \ \sum_{p \le N} \log^2 p. \end{aligned} \tag{17}$$

Using $\sum_{p \le N} \log p = N \ + \ \text{small}$ and Partial Summation, we can show

$$\sum_{p \le N} \log^p \;=\; N \log N \;+\; \text{smaller}. \tag{18}$$

Thus,

$$\int_0^1 |F_N(x)|^2 \;=\; N \log N \;+\; \text{smaller}. \tag{19}$$

Thus, taking square-roots, we see on average $|\Big(F_N(x)\Big)|^2$ is $N \log N$, significantly smaller than the maximum possible value (N^2). Thus, we see we are getting almost square-root cancellation on average. □

2.4 Definition of the Major and Minor Arcs

We split the unit interval $[0, 1)$ into two disjoint parts, the Major and the Minor arcs.

Roughly, the Major arcs will be a union of very small intervals centered at rationals with small denominator (relative to N). Near these rationals, we will be able to approximate $F_N(x)$ very well, and $F_N(x)$ will be of size N.

The minor arcs will be the rest of $[0, 1)$; we will show that $F_N(x)$ is significantly smaller than N here.

2.4.1 Major Arcs

Let $B > 0$, and let $Q = (\log N)^B \ll N$.

For each $q \in \{1, 2, \ldots, Q\}$ and $a \in \{1, 2, \ldots, q\}$ with a and q relatively prime, consider the set

$$\mathcal{M}_{a,q} \;=\; \Big\{ x \in [0, 1) : \; \Big| x - \frac{a}{q} \Big| \;<\; \frac{Q}{N} \Big\}. \tag{20}$$

We also add in one interval centered at either 0 or 1, ie, the "interval" (or wrapped-around interval)

$$\left[0, \frac{Q}{N}\right] \;\cup\; \left[1 - \frac{Q}{N}, 1\right]. \tag{21}$$

Exercise 2.8. *Show, if N is large, that the major arcs $\mathcal{M}_{a,q}$ are disjoint for $q \le Q$ and $a \le q$, a and q relatively prime.*

We define the Major Arcs to be the union of each arc $\mathcal{M}_{a,q}$:

$$\mathcal{M} \ = \ \bigcup_{q=1}^{Q} \bigcup_{\substack{a=1 \\ (a,q)=1}} \mathcal{M}_{a,q}, \tag{22}$$

where (a, q) is the greatest common divisor of a and q.

Exercise 2.9. *Show* $|\mathcal{M}| < \frac{2Q^3}{N}$*. As* $Q = \log^B N$*, this implies as* $N \to \infty$*, the major arcs are zero percent of the unit interval.*

2.4.2 Minor Arcs

The Minor Arcs, m, are whatever is *not* in the Major Arcs. Thus,

$$\mathrm{m} \ = \ [0, 1) - \mathcal{M}. \tag{23}$$

Clearly, as $N \to \infty$, almost all of $[0, 1)$ is in the Minor Arcs.

3 Contributions from the Major and Minor Arcs

3.1 Contribution from the Minor Arcs

We bound the contribution from the minor arcs to $r_N(m)$:

$$\begin{aligned}
\Big| \int_{\mathrm{m}} F_N^3(x) e(-mx) dx \Big| \ &\leq \ \int_{\mathrm{m}} |F_N(x)|^3 dx \\
&\leq \ \Big(\max_{x \in \mathrm{m}} |F_N(x)| \Big) \int_{\mathrm{m}} |F_N(x)|^2 dx \\
&\leq \ \Big(\max_{x \in \mathrm{m}} |F_N(x)| \Big) \int_0^1 F_N(x) F_N(-x) dx \\
&\leq \ \Big(\max_{x \in \mathrm{m}} |F_N(x)| \Big) N \log N. \qquad (24)
\end{aligned}$$

As the minor arcs are most of the unit interval, replacing $\int_{\mathrm{m}}$ with $\int_0^1$ doesn't introduce much of an over-estimation.

In order for the Circle Method to succeed, we need a non-trivial, good bound for

$$\max_{x \in \mathfrak{m}} |F_N(x)| \tag{25}$$

This is where most of the difficulty arises, showing that there is good cancellation in $F_N(x)$ if we stay away from rationals with small denominator.

We will show that the contribution to the major arcs is

$$\mathfrak{S}(N)\frac{N^2}{2} \ + \ \text{smaller}, \tag{26}$$

where $\exists c_1, c_2 > 0$ such that, for all N, $c_1 < \mathfrak{S}(N) < c_2$.

Thus, we need the estimate that

$$\max_{x \in \mathfrak{m}} |F_N(x)| \ \leq \ \frac{N}{\log^{1+\epsilon} N}. \tag{27}$$

Relative to the average size of $|F_N(x)|^2$, this is significantly smaller; however, as we are showing that the maximum value of $|F_N(x)|$ is bounded, this is a significantly more delicate question. We know such a bound cannot be true for all $x \in [0,1)$ (see below, and not that $F_N(0) = N$). The hope is that if x is not near a rational with small denominator, we will get moderate cancellation.

While this is very reasonable to expect, it is not easy to prove.

3.2 Contribution from the Major Arcs

Fix a $q \leq Q$ and an $a \leq q$ with a and q relatively prime. We evaluate $F\left(\frac{a}{q}\right)$.

$$\begin{aligned}
F\left(\frac{a}{q}\right) &= \sum_{p\leq N} \log p \cdot e^{2\pi i p \frac{a}{q}} \\
&= \sum_{r=1}^{q} \sum_{\substack{p\equiv r(q)\\ p\leq N}} \log p \cdot e^{2\pi i \frac{ap}{q}} \\
&= \sum_{r=1}^{q} \sum_{\substack{p\equiv r(q)\\ p\leq N}} \log p \cdot e^{2\pi i \frac{ar}{q}} \\
&= \sum_{r=1}^{q} e^{2\pi i \frac{ar}{q}} \sum_{\substack{p\equiv r(q)\\ p\leq N}} \log p \qquad (28)
\end{aligned}$$

Note the beauty of the above. The dependence on p in the original sums is very weak – there is a $\log p$ factor, and there is $e\left(\frac{ap}{q}\right)$. In the exponential, we only need to know p mod q. Now, p runs from 2 to N, and q is at most $\log^B N$. Thus, in general $p \gg q$.

We use the Siegel-Walfisz Theorem. We first remark that we may assume r and q are relatively prime. Why? If $p \equiv r$ mod q, this means $p = \alpha q + r$ for some $\alpha \in \mathbb{N}$. If r and q have a common factor, there can be at most one prime p (namely r) such that $p \equiv r$ mod q, and this can easily be shown to give a negligible contribution.

For any $C > 0$

$$\sum_{\substack{p\equiv r(q)\\ p\leq N}} \log p = \frac{N}{\phi(q)} + O\left(\frac{N}{\log^C N}\right). \qquad (29)$$

Now, as $\phi(q)$ is at most q which is at most $\log^B N$, we see that the main term is significantly greater than the error term (choose C much greater than B).

Note the Siegel-Walfisz Theorem would be useless if $q \approx N^\epsilon$. Then the main term would be like $N^{1-\epsilon}$, which would be smaller than the error term.

This is one reason why, in constructing the major arcs, we take the denominators to be small.

Thus, we find

$$
\begin{aligned}
F\left(\frac{a}{q}\right) &= \sum_{\substack{r=1 \\ (r,q)=1}}^{q} e^{2\pi i \frac{ar}{q}} \frac{N}{\phi(q)} + \text{smaller} \\
&= \frac{N}{\phi(q)} \sum_{\substack{r=1 \\ (r,q)=1}}^{q} e^{2\pi i \frac{ar}{q}}. \qquad (30)
\end{aligned}
$$

We merely sketch what happens now.

First, one shows that for $x \in \mathcal{M}_{a,q}$ that $F_N(x)$ is very close to $F\left(\frac{a}{q}\right)$. This is a standard analysis (Taylor Series Expansion – the constant term is a good approximation if you are sufficiently close).

Thus, as the major arcs are distinct,

$$
\int_{\mathcal{M}} F_N^3(x) e(-mx) dx = \sum_{q=1}^{Q} \sum_{\substack{a=1 \\ (a,q)=1}} \int_{\mathcal{M}_{a,q}} F_N^3(x) e(-mx) dx. \qquad (31)
$$

We can approximate $F_N^3(x)$ by $F\left(\frac{a}{q}\right)$; integrating a constant gives the constant times the length of the interval. Each of the major arcs has length $\frac{2Q^3}{N}$. Thus we find that, up to a smaller correction term, the contribution from the Major Arcs is

$$
\begin{aligned}
\int_{\mathcal{M}} F_N^3(x) e(-mx) dx &= \frac{2Q^3}{N} \sum_{q=1}^{Q} \sum_{\substack{a=1 \\ (a,q)=1}} \left(\frac{N}{\phi(q)} \sum_{\substack{r=1 \\ (r,q)=1}}^{q} e^{2\pi i \frac{ar}{q}} \right)^3 e\Big(\frac{-2\pi i m a}{q}\Big) \\
&= N^2 \cdot 2Q^3 \sum_{q=1}^{Q} \frac{1}{\phi(q)^3} \sum_{\substack{a=1 \\ (a,q)=1}} \left(\sum_{\substack{r=1 \\ (r,q)=1}}^{q} e^{2\pi i \frac{ar}{q}} \right)^3 e\Big(\frac{-2\pi i m a}{q}\Big).
\end{aligned}
$$

(32)

To complete the proof, we need to show that what is multiplying N^2 is non-negative, and not too small.

We will leave this for another day, as it is getting quite late here.

4 Why Goldbach is Hard

Using

$$F_N(x) \;=\; \sum_{p \leq N} \log p \cdot e^{2\pi i p x}, \tag{33}$$

we find we must study

$$\int_0^1 F_N^s(x)dx, \tag{34}$$

where $s = 3$ if we are looking at $p_1 + p_2 + p_3 = 2n + 1$ and $s = 2$ if we are looking at $p_1 + p_2 = 2n$. Why does the circle method work for $s = 3$ but fail for $s = 2$?

4.1 $s = 3$ Sketch

Let us recall *briefly* the $s = 3$ case. Near rationals $\frac{a}{q}$ with *small* denominator (*small* means $q \leq \log^B N$), we can evaluate $F_N(\frac{a}{q})$. Using Taylor, if x is very close to $\frac{a}{q}$, we expect $F_N(x)$ to be close to $F_N(\frac{a}{q})$.

The Major Arcs have size $\frac{\log^B N}{N}$. As $F_N(x)$ is around N near such rationals, we expect the integral of $F_N^3(x)e(-mx)$ to be N^2 times a power of $\log N$. Doing a careful analysis of the singular series shows that the contribution is actually $\mathfrak{S}(N)N^2$, where there exist constants independent of N such that $0 < c_1 < \mathfrak{S}(N) < c_2 < \infty$.

A direct calculation shows that

$$\int_0^1 |F_N(x)|^2 dx \;=\; \int_0^1 F_N(x)F_N(-x)dx \;=\; N. \tag{35}$$

Thus, if m denotes the minor arcs,

$$\begin{aligned} \Big| \int_{\mathrm{m}} F_N^3(x)e(-mx)dx \Big| \;&\leq\; \max_{x \in \mathrm{m}} |F_N(x)| \int_0^1 |F_N(x)|^2 dx \\ &\leq\; N \max_{x \in \mathrm{m}} |F_N(x)|. \end{aligned} \tag{36}$$

As the major arcs contribute $\mathfrak{S}(N)N^2$, we need to show

$$\max_{x\in\mathfrak{m}} |F_N(x)| \ \ll \ \frac{N}{\log^D N}. \tag{37}$$

Actually, we just need to show the above is $\ll o(N)$. This is the main difficulty – the trivial bound is $|F_N(x)| \le N$. As $F_N(0) = N$ plus lower order terms, we cannot do better in general.

Exercise 4.1. *Show $F_N(\frac{1}{2}) = N - 1$ plus lower order terms.*

The key observation is that, if we stay away from rationals with small denominator, we can prove there is cancellation in $F_N(x)$. While we don't go into details here (see, for example, Nathanson's Additive Number Theory: The Classical Bases, Chapter 7), the savings we obtain is small. We show

$$\max_{x\in\mathfrak{m}} |F_N(x)| \ \ll \ \frac{N}{\log^D N}. \tag{38}$$

Note that Equation 35 gives us significantly better cancellation on average, telling us that $|F_N(x)|^2$ is usually of size N.

Thus, it is our dream to be so lucky as to see $\left|\int_I |F_N(x)|^2 dx\right|$ for any $I \subset [0,1)$, as we can evaluate this extremely well.

4.2 $s = 2$ Sketch

What goes wrong when $s = 2$? As a first approximation, if $s = 3$ has the Major Arcs contributing a constant times N^2 (and $F_N(x)$ was of size N on the Major Arcs), one might guess that the Major Arcs for $s = 2$ will contribute a constant times N.

How should we estimate the contribution from the Minor Arcs? We have $F_N^2(x)$. If we just throw in absolute values we get

$$\left|\int_{\mathfrak{m}} F_N^2(x) e(-mx) dx\right| \ \le \ \int_0^1 |F_N(x)|^2 dx \ = \ N. \tag{39}$$

Note, unfortunately, that this is the same size as the expected contribution from the Major Arcs!

We could try pulling a $\max_{x\in\mathfrak{m}} |F_N(x)|$ outside the integral, and hope to get a good savings. The problem is this leaves us with $\int_{\mathfrak{m}} |F_N(x)| dx$.

Recall

Lemma 4.2.

$$\int_0^1 |f(x)g(x)|dx \;\le\; \Big(\int_0^1 |f(x)|^2dx\Big)^{\frac{1}{2}} \cdot \Big(\int_0^1 |g(x)|^2dx\Big)^{\frac{1}{2}}. \tag{40}$$

For a proof, see Lemma A.1.

Thus,

$$\begin{aligned}\Big|\int_{\mathrm{m}} F_N^2(x)e(-mx)dx\Big| &\le \max_{x\in \mathrm{m}} |F_N(x)| \int_0^1 |F_N(x)|dx \\ &\le \max_{x\in \mathrm{m}} |F_N(x)| \Big(\int_0^1 |F_N(x)|^2dx\Big)^{\frac{1}{2}} \cdot \Big(\int_0^1 1^2dx\Big)^{\frac{1}{2}} \\ &\le \max_{x\in \mathrm{m}} |F_N(x)| \cdot N^{\frac{1}{2}} \cdot 1. \qquad (41)\end{aligned}$$

As the Major Arcs contribute something of size N, we would need

$$\max_{x\in \mathrm{m}} |F_N(x)| \;\ll\; o(\sqrt{N}). \tag{42}$$

There is almost no chance of such cancellation. We know

$$\int_0^1 |F_N(x)|^2dx \;=\; N \text{ plus lower order terms.} \tag{43}$$

Thus, the average size of $|F_N(x)|$ is N, so we expect $|F_N(x)|$ to be about $\sqrt{N}$. To get $o(N)$ would be unbelievably good fortune!

While the above sketch shows the Circle Method is not, at present, powerful enough to handle the Minor Arc contributions, all is not lost. The quantity we *need* to bound is

$$\Big|\int_{\mathrm{m}} F_N^2(x)e(-mx)dx\Big|. \tag{44}$$

However, we have instead been studying

$$\int_{\mathrm{m}} |F_N(x)|^2dx \tag{45}$$

and

$$\max_{x\in \mathrm{m}} |F_N(x)| \int_0^1 |F_N(x)|dx. \tag{46}$$

Thus, we are ignoring the probable oscillation / cancellation in the integral $\int F_N(x)e(-mx)dx$. It is *this cancellation* that will lead to the Minor Arcs contributing significantly less than the Major Arcs.

However, showing there is cancellation in the above integral is very difficult. It is a lot easier to work with absolute values.

A Cauchy-Schwartz Inequality

Lemma A.1. *[Cauchy-Schwarz]*

$$\int_0^1 |f(x)g(x)|dx \ \leq \ \left(\int_0^1 |f(x)|^2 dx\right)^{\frac{1}{2}} \cdot \left(\int_0^1 |g(x)|^2 dx\right)^{\frac{1}{2}}. \tag{47}$$

For notational simplicity, assume f and g are real-valued, positive functions. Working with $|f|$ and $|g|$ we see there is no harm in the above.

Let

$$h(x) = \ f(x) + \lambda g(x), \quad \lambda \ = \ -\frac{\int_0^1 f(x)g(x)dx}{\int_0^1 g(x)^2 dx} \tag{48}$$

As $\int_0^1 h(x)^2 dx \geq 0$, we have

$$
\begin{aligned}
0 &\leq \int_0^1 \Big(f(x)+\lambda g(x)\Big)^2 dx \\
&= \int_0^1 f(x)^2dx \ + \ 2\lambda \int_0^1 f(x)g(x)dx \ + \ \lambda^2 \int_0^1 g(x)^2dx \\
&= \int_0^1 f(x)^2dx \ - \ 2\frac{\Big(\int_0^1 f(x)g(x)dx\Big)^2}{\int_0^1 g(x)^2dx} \ + \ \frac{\Big(\int_0^1 f(x)g(x)dx\Big)^2}{\int_0^1 g(x)^2dx} \\
&= \int_0^1 f(x)^2dx \ - \ \frac{\Big(\int_0^1 f(x)g(x)dx\Big)^2}{\int_0^1 g(x)^2dx} \\
\frac{\Big(\int_0^1 f(x)g(x)dx\Big)^2}{\int_0^1 g(x)^2dx} &\leq \int_0^1 f(x)^2dx \\
\Big(\int_0^1 f(x)g(x)dx\Big)^2 &\leq \int_0^1 f(x)^2dx \cdot \int_0^1 g(x)^2dx \\
\int_0^1 f(x)g(x)dx &\leq \Big(\int_0^1 f(x)^2dx\Big)^{\frac{1}{2}} \cdot \Big(\int_0^1 g(x)^2dx\Big)^{\frac{1}{2}}.
\end{aligned}
\tag{49}
$$

Again, for general f and g, replace $f(x)$ with $|f(x)|$ and $g(x)$ with $|g(x)|$ above. Note there is nothing special about $\int_0^1$. □

The Cauchy-Schwarz Inequality is often useful when $g(x) = 1$. In this special case, it is important that we integrate over a finite interval.

Exercise A.2. *For what f and g is the Cauchy-Schwarz Inequality an equality?*

B Partial Summation

Lemma B.1 (Partial Summation: Discrete Version). *Let $A_N = \sum_{n=1}^{N} a_n$. then*

$$\sum_{n=M}^{N} a_n b_n = A_N b_N - A_{M-1} b_M + \sum_{n=M}^{N-1} A_n (b_n - b_{n+1}) \tag{50}$$

Proof. Since $A_n - A_{n_1} = a_n$,

$$\begin{aligned}
\sum_{n=M}^{N} a_n b_n &= \sum_{n=M}^{N} (A_n - A_{n-1}) b_n \\
&= (A_N - A_{N-1}) b_N + (A_{N-1} - A_{N-2}) b_{N-1} + \cdots + (A_M - A_{M-1}) b_M \\
&= A_N b_N + (-A_{N-1} b_N + A_{N-1} b_{N-1}) + \cdots + (-A_M b_{M+1} + A_M b_M) - a_{M-1} b_M \\
&= A_N b_N - a_{M_1} b_M + \sum_{n=M}^{N-1} A_n (b_n - b_{n+1}). \qquad (51)
\end{aligned}$$

□

Lemma B.2 (Abel's Summation Formula - Integral Version). *Let $h(x)$ be a continuously differentiable function. Let $A(x) = \sum_{n \le x} a_n$. Then*

$$\sum_{n \le x} a_n h(n) = A(x) h(x) - \int_1^x A(u) h'(u) du \qquad (52)$$

See, for example, W. Rudin, *Principles of Mathematical Analysis*, page 70.

Partial Summation allows us to take knowledge of one quantity and convert that to knowledge of another.

For example, suppose we know that

$$\sum_{p \le x} \log p \;=\; x + O(x^{\frac{1}{2}+\epsilon}). \qquad (53)$$

We use this to glean information about $\sum_{p \le x} 1$.

Define

$$h(n) = \frac{1}{\log n} \quad \text{and} \quad a_n = \begin{cases} \log n & \text{if } n \text{ is prime} \\ 0 & \text{otherwise.} \end{cases} \qquad (54)$$

Applying partial summation to $\sum_{p \le x} a_n h(n)$ will give us knowledge about $\sum_{p \le x} 1$. Note as long as $h(n) = \frac{1}{\log n}$ for n prime, it doesn't matter how we define $h(n)$ elsewhere; however, to use the integral version of Partial Summation, we need h to be a differentiable function.

Thus

$$\begin{aligned}\sum_{p\le x} 1 &= \sum_{p\le x} a_n h(n) \\ &= \left(x + O(x^{\frac{1}{2}+\epsilon})\right)\frac{1}{\log x} - \int_2^x \left(u + O(u^{\frac{1}{2}+\epsilon})\right)h'(u)du. \qquad (55)\end{aligned}$$

The main term $(A(x)h(x))$ equals $\frac{x}{\log x}$ plus a significantly smaller error.

We now calculate the integral, noting $h'(u) = -\frac{1}{u\log^2 u}$. The error piece in the integral gives a constant multiple of

$$\int_2^x \frac{u^{\frac{1}{2}+\epsilon}}{u\log^2 u} du. \qquad (56)$$

As $\frac{1}{\log^2 u} \le \frac{1}{\log^2 2}$ for $2 \le u \le x$, the integral is bounded by

$$\frac{1}{\log^2 2}\int_2^x u^{-\frac{1}{2}+\epsilon} < \frac{1}{\log^2 2}\frac{1}{\frac{1}{2}+\epsilon}x^{\frac{1}{2}+\epsilon}, \qquad (57)$$

which is significantly less than $A(x)h(x) = \frac{x}{\log x}$.

We now need to handle the other integral:

$$\int_2^x \frac{u}{u\log^2 u} du = \int_2^x \frac{1}{\log^2 u} du. \qquad (58)$$

The obvious approximation to try is $\frac{1}{\log^2 u} \le \frac{1}{\log^2 2}$. Unfortunately, plugging this in bounds the integral by $\frac{x}{\log^2 2}$. This is larger than the expected main term, $A(x)h(x)$!

As a rule of thumb, whenever you are trying to bound something, try the simplest, most trivial bounds first. Only if they fail should you try to be clever.

Here, we need to be clever, as we are bounding the integral by something larger than the observed terms.

We split the integral into two pieces:

$$\int_2^x = \int_2^{\sqrt{x}} + \int_{\sqrt{x}}^x \qquad (59)$$

For the first piece, we use the trivial bound for $\frac{1}{\log^2 u}$. Note the interval has length $\sqrt{x} - 2 < \sqrt{x}$. Thus, the first piece contributes at most $\frac{x^{\frac{1}{2}}}{\log^2 2}$, significantly less than $A(x)h(x)$.

The reason trivial bounds failed for the entire integral is the length was too large (of size x); there wasn't enough decay in the function.

The advantage of splitting the integral in two is that in the second piece, even though most of the length of the original interval is here (it is of length $x - \sqrt{x} \approx x$), the function $\frac{1}{\log^2 u}$ is small here. Instead of bounding it by a constant, we now bound it by substituting in the smallest value of u on this interval, $\sqrt{x}$. Thus, the contribution from this integral is at most $\frac{x-\sqrt{x}}{\log^2 \sqrt{x}} < \frac{4x}{\log^2 x}$. Note that this is significantly less than the main term $A(x)h(x) = \frac{x}{\log x}$.