Aspectos Ergódicos da Teoria dos Números

Publicações Matemáticas

Aspectos Ergódicos da Teoria dos Números

Alexander Arbieto
UFRJ

Carlos Matheus
IMPA

Carlos Gustavo Moreira
IMPA

impa

26º Colóquio Brasileiro de Matemática

Impresso no Brasil / Printed in Brazil

Capa: Noni Geiger / Sérgio R. Vaz

26º Colóquio Brasileiro de Matemática

- **Aspectos Ergódicos da Teoria dos Números - Alexander Arbieto, Carlos Matheus e Carlos Gustavo Moreira**
- Componentes Irredutíveis dos Espaços de Folheações - Alcides Lins Neto
- Elliptic Regularity and Free Boundary Problems: an Introduction - Eduardo V. Teixeira
- Hiperbolicidade, Estabilidade e Caos em Dimensão Um - Flavio Abdenur e Luiz Felipe Nobili França
- Introduction to Generalized Complex Geometry - Gil R. Cavalcanti
- Introduction to Tropical Geometry - Grigory Mikhalkin
- Introdução aos Algoritmos Randomizados - Celina de Figueiredo, Guilherme da Fonseca, Manoel Lemos e Vinicius de Sá
- Mathematical Aspects of Quantum Field Theory - Edson de Faria and Welington de Melo
- Métodos Estatísticos Não-Paramétricos e suas Aplicações - Aluisio Pinheiro e Hildete P. Pinheiro
- Moduli Spaces of Curves - Enrico Arbarello
- Noções de Informação Quântica - Marcelo O. Terra Cunha
- Three Dimensional Flows - Vítor Araújo e Maria José Pacifico
- Tópicos de Corpos Finitos com Aplicações em Criptografia e Teoria de Códigos - Ariane Masuda e Daniel Panario
- Tópicos Introdutórios à Análise Complexa Aplicada - André Nachbin e Ailín Ruiz de Zárate
- Uma Introdução à Mecânica Celeste - Sérgio B. Volchan
- Uma Introdução à Teoria Econômica dos Jogos - Humberto Bortolossi, Gilmar Garbugio e Brígida Sartini
- Uma Introdução aos Sistemas Dinâmicos via Frações Contínuas - Lorenzo J. Díaz e Danielle de Rezende Jorge

ISBN: 978-85-244-0250-0

Distribuição: IMPA
Estrada Dona Castorina, 110
22460-320 Rio de Janeiro, RJ
E-mail: ddic@impa.br
http://www.impa.br

Dedicamos este livro ao Ramiro Mendoza Sánchez,
à Aline Gomes Cerqueira,
ao Carlos de Araujo Moreira Neto e
à Raquel Tavares Scarpelli.

Prefácio

Problemas que envolvem simultaneamente as estruturas aditiva e multiplicativa dos números inteiros, em particular problemas aditivos sobre números primos, costumam ser extremamente difíceis, apesar de muitas vezes terem enunciados bastante simples. Não se sabe por exemplo se há infinitos pares de primos gêmeos, i.e., pares de primos cuja diferença é 2. Também continua em aberto a famosa conjectura de Goldbach: todo inteiro par maior ou igual a 4 é soma de dois primos.

Outra conjectura clássica sobre primos que estava em aberto há muito tempo é a de que existem progressões aritméticas arbitrariamente longas formadas exclusivamente por primos. A maior dessas progressões conhecida atualmente é $468395662504823 + k \cdot 45872132836530, 0 \leq k \leq 23$, formada por 24 primos, descoberta em 18 de janeiro de 2007 por Jaroslaw Wroblewski. Esta conjectura foi finalmente demonstrada por Ben Green e Terence Tao em 2004. Tao ganhou uma medalha Fields em 2006, principalmente por causa deste trabalho.

O objetivo principal deste texto é expor o trabalho de Green e Tao da forma mais auto-contida possível. Sua demonstração utiliza o famoso Teorema de Szemerédi, segundo o qual qualquer conjunto de inteiros positivos com densidade (superior) positiva contém progressões aritméticas arbitrariamente longas. O trabalho de Green e Tao usa ainda idéias de teoria ergódica, introduzidas por Furstenberg para dar uma prova alternativa do Teorema de Szemerédi, além de técnicas introduzidas por Gowers para dar ainda outra demonstração deste Teorema de Szemerédi.

No capítulo 1 apresentaremos diversos resultados sobre números primos, incluindo a demonstração do Teorema dos Números Primos, sobre sua distribuição assintótica, durante a qual faremos estimativas sobre a função ζ de Riemann que serão usadas na prova do Teorema de Green e Tao. Discutiremos também resultados ligados ao Teorema de Szemerédi e a prova ergódica de Furstenberg.

No capítulo 2 apresentaremos a demonstração do Teorema de Green e Tao que generaliza o Teorema de Szemerédi via a introdução das *medidas pseudo aleatórias*. Neste capítulo aparecem as técnicas ergódicas e as técnicas de Gowers que mencionamos.

No capítulo 3 provamos que existem medidas pseudo-aleatórias em relação às quais os primos têm medida positiva, o que, pelos resultados do capítulo 2, permite concluir a existência de progressões aritméticas arbitrariamente longas formadas por primos.

Apesar de sofisticada, a prova do Teorema de Green-Tao não requer muitos pré-requisitos que não estejam contidos neste texto (em particular não usaremos diretamente resultados de teoria ergódica nem de teoria analítica dos números que não estejam demonstrados nestas notas; por outro lado, alguma experiência prévia com esses assuntos pode ajudar a compreender muitas das idéias da prova).

Sumário

Capítulo 1

Propriedades aditivas dos números primos

1.1 Introdução

Um dos conceitos numéricos mais antigos é a noção de número primo. Por definição um número p é primo se ele é divisível somente por 1 e por ele mesmo.

Os números primos aparecem em diversos resultados elementares da teoria dos números, como o teorema de decomposição única em fatores primos ou como os números tais que $\mathbb{Z}/p\mathbb{Z}$ é um corpo.

Obviamente como a definição de número primo é de caráter *multiplicativo*, podemos extrair diversas propriedades multiplicativas elementares. Por exemplo, o produto de dois primos não é primo. Ou mesmo, não existem progressões geométricas de comprimento maior ou igual à 3 formadas somente por primos.

Por outro lado, ao levantarmos perguntas de caráter *aditivo* podemos nos deparar com algumas surpresas. Por exemplo, a soma de dois números primos é primo? A resposta é: depende. Por exemplo 2+3=5 é primo, 2+5=7 é primo, mas 3+5=8 não é primo e nem 7+2=9. Por outro lado o postulado de Bertrand diz que para todo natural N existe um primo entre N e $2N$. Vê-se então que a seguinte pergunta merece pelo menos um pouco de reflexão:

Existem progressões aritméticas de comprimento maior ou igual à 3 formadas somente por primos? E quantas existem uma vez que o comprimento for fixado?

Veremos nestas notas como tal pergunta foi respondida por Ben Green e Terence Tao. Mas, antes disso, iremos passear pelo mundo dos números primos, vendo soluções parciais a esta pergunta e analisando outras questões de caráter aditivo envolvendo os números primos.

1.2 Problemas clássicos sobre propriedades aditivas de Números Primos

1.2.1 A conjectura dos primos gêmeos

Observando o exemplo da introdução, sabemos que nem sempre a soma de um número primo com 2 é primo; mas, será que existem infinitos primos com essa propriedade? Dizemos que p e p+2 são *primos gêmeos* se ambos são primos. Exemplos de primos gêmeos são: (3 e 5), (5 e 7), (11 e 13), (17 e 19), (29 e 31), (41 e 43). Um dos problemas em aberto mais famosos da teoria dos números é a *conjectura dos primos gêmeos*:

Existem infinitos primos gêmeos?

Um resultado importante devido a Brun [2] mostra que mesmo que existam infinitos primos gêmeos, eles se tornam muito escassos quando olhamos para números muito grandes, o que torna a conjectura mais difícil. De fato o teorema de Brun diz que a série dos inversos dos primos gêmeos ímpares converge (para um número conhecido como a constante de Brun):

$$(\frac{1}{3}+\frac{1}{5})+(\frac{1}{5}+\frac{1}{7})+(\frac{1}{11}+\frac{1}{13})+(\frac{1}{17}+\frac{1}{19})+...<+\infty.$$

Mais tarde reformularemos a conjectura dos primos gêmeos em uma linguagem mais analítica.

1.2.2 A conjectura de Goldbach

Em uma carta a Euler, em 1742, Goldbach perguntava se todo número maior que 2 é soma de 3 primos. Goldbach assumia que 1 era primo, o que não é mais usado. Portanto uma conjectura equivalente é a famosa conjectura de Goldbach é:

Todo inteiro par $n \geq 4$ pode ser escrito como soma de dois primos?

Mesmo sendo fácil de enunciar, a conjectura de Goldbach ainda é um dos maiores desafios da teoria dos números. Diversos resultados parciais foram obtidos, mas nenhuma das provas parece se estender a uma demonstração da conjectura de Goldbach.

Por exemplo, Schnirelman [8] mostrou que todo número primo pode ser escrito como uma soma de primos, porém o número de parcelas é maior que 300000, um pouco longe de 2, não?

Outra conjectura relacionada é chamada de conjectura fraca de Goldbach:

Todo número ímpar $n \geq 9$ pode ser escrito como soma de 3 primos?

Com respeito a este problema, temos o famoso teorema de Vinogradov [15], onde ele resolve a conjectura fraca de Goldbach para números ímpares suficientemente grandes (maiores que $3^{3^{15}}$).

Outro resultado interessante é o teorema de Chen [3], onde ele mostra que um número par suficientemente grande é soma de um primo com um quase-primo (um número com no máximo 2 fatores primos).

Uma versão mais forte da conjectura fraca de Goldbach é conhecida como a conjectura de Levy:

Todo número ímpar $n \geq 7$ pode ser escrito como soma de um primo mais o dobro de outro primo?

Mais adiante, reformularemos estas conjecturas de maneira analítica.

1.2.3 Primeiros Resultados sobre Progressões Aritméticas e Números Primos

Um dos resultados mais clássicos neste assunto é o teorema de Dirichlet que diz:

Se a e b são primos entre si então a progressão aritmética $a + nb$ contém infinitos primos.

A prova deste resultado utiliza o conceito de *L-séries* (uma definição mais avançada) devido a Dirichlet. No apêndice será dado um esboço da prova em um caso particular.

O teorema de Dirichlet não diz que a progressão aritmética é formada *inteiramente* de primos. Uma pergunta natural é se existe uma progressão aritmética de tamanho infinito formada somente de primos. A resposta é negativa segundo o teorema de Lagrange-Waring:

Considere uma progressão aritmética formada somente de primos de comprimento k e de razão d. Então necessariamente d é divisível por todos os primos menores que k. Em particular não existem progressões aritméticas de comprimento infinito formadas somente de primos.

1.3 Progressões Aritméticas em certos subconjuntos de $\mathbb{Z}$

A questão da existência de progressões aritméticas de tamanho finito formadas de primos pode ser estendida da seguinte maneira:

Seja $A \subset \mathbb{Z}$. Existem progressões aritméticas de comprimento arbitrariamente grande formadas somente por números que pertencem a A?

De certa forma, veremos que o conjunto P de números primos é muito "magro". Podemos então tentar atacar o problema primeiramente para conjuntos A "gordos", onde as chances de se encontrar progressões aritméticas sao maiores, e tentar adaptar os métodos de prova para o caso de conjuntos "magros". Obviamente um problema central é a definição do que é um conjunto "magro" e o que é um conjunto "gordo". Nesta seção veremos certos resultados nesta direção. Observe que não aplicaremos a ordem cronológica na exposição dos resultados.

1.3.1 O teorema de Van der Waerden

Suponha que você possui uma quantidade finita, digamos k, de cores e use-as para pintar os números inteiros. Então, você obtém k subconjuntos disjuntos que formam uma partição dos inteiros. O teorema de Van der Waerden diz que:

Pelo menos um destes subconjuntos é tão "gordo" que possui progressões aritméticas de comprimento arbitrariamente grande.

Em particular, se tomamos duas cores e pintamos os primos de uma cor e os não-primos de outra, obtemos:

O conjunto de números primos ou o conjunto de números não-primos possuem progressões aritméticas de comprimento arbitrário.

Mais adiante veremos provas do teorema de Van der Waerden.

1.3.2 Conjuntos com Densidade Positiva e o Teorema de Szemerédi

Obviamente, o conjunto dos números pares possuem progressões aritméticas de comprimento arbitrário (com razão 2, por exemplo). Observe que num intervalo $[1, N] := \{n \in \mathbb{Z}; 1 \leq n \leq N\}$ essencialmente os pares ocupam $1/2$ deste conjunto. Da mesma maneira, os números ímpares também tem essa propriedade e possuem progressões aritméticas de comprimento arbitrário. Mais geralmente, escolhido um número k qualquer se você olha para o conjunto de múltiplos de k, este conjunto essencialmente ocupa $1/k$ de $[1, N]$ e possui progressões aritméticas de comprimento arbitrário.

Com base nisto, podemos tentar dizer que um conjunto é "gordo" se ele ocupa uma fração positiva do intervalo $[1, N]$. Por outro lado, como queremos progressões de comprimento grande, iremos pedir que essa fração seja vista assintoticamente.

Definição 1.3.1. *Seja $A \subset \mathbb{N}$ a densidade de A é:*

$$d(A) = \limsup_{N \to \infty} \frac{|[1, N] \cap A|}{N}.$$

Aqui, dado $B \subset \mathbb{N}$, denotamos por $|B|$ a cardinalidade de B.

Obviamente a definição se estende naturalmente para subconjuntos dos inteiros. O primeiro teorema que lida com conjuntos "gordos" ou melhor com densidade positiva é o teorema de Roth [7] de 1956:

Se $A \subset \mathbb{Z}$ tem densidade positiva então A possui infinitas progressões aritméticas de comprimento 3.

O problema da existência de progressões aritméticas de comprimento arbitrário somente foi resolvido graças aos trabalhos de Szemerédi [9] em 1975:

Teorema 1.3.1 (Szemerédi)**.** *Se $A \subset \mathbb{Z}$ tem densidade positiva então A possui infinitas progressões aritméticas de comprimento arbitrariamente grande.*

Adaptações da prova do teorema de Szemerédi serão objeto de estudo nos capítulos posteriores, pois veremos a seguir que não se pode aplicar o teorema nesta forma ao conjunto dos números primos. Nas seções seguintes, daremos provas do teorema de Szemerédi.

1.3.3 O teorema dos Números Primos e Progressões Aritméticas formadas por Primos

O motivo pelo qual não podemos aplicar o teorema de Szemerédi ao conjunto de números primos se deve ao famoso teorema dos números primos:

Teorema 1.3.2 (O Teorema dos Números Primos)**.** *Vale a seguinte estimativa assintótica:*

$$\frac{|P \cap [1, N]|}{N} = \frac{1}{\log N} + o(1).$$

Aqui P é o conjunto de números primos e $o(1)$ é uma quantidade que vai a zero quando $N \to \infty$. Em particular $d(P) = 0$.

Mesmo que os primos tenham densidade zero, a existência de infinitas progressões aritméticas de comprimento 3 formada de primos foi obtida em 1939 por Van der Corput (antes do teorema de Roth):

Existem infinitas progressões artiméticas de comprimento 3 formadas somente de primos.

Finalmente, em 2004, Ben Green e Terence Tao [5] obtiveram o resultado geral. Tal teorema é objeto central de estudo deste livro:

Teorema 1.3.3 (Green-Tao)**.** *Existem infinitas progressões aritméticas de comprimento arbitrário formadas somente de primos.*

1.3.4 A conjectura de Erdös-Turán

Sabe-se que a série $\sum \frac{1}{n^2}$ converge, porém, em 1737, Euler mostrou que a série dos inversos dos primos diverge:

$$\sum_{p \text{ primo}} \frac{1}{p} = +\infty.$$

Isto mostra que os números primos não são tão esparsos quanto os quadrados de números naturais.

A conjectura de Erdös-Turán diz que conjuntos com tal propriedade devem conter progressões aritméticas de comprimento arbitrário. O teorema de Green-Tao é portanto um caso particular desta conjectura:

Conjectura 1 (Erdös-Turán)**.** Seja $A \subset \mathbb{N}$ tal que:

$$\sum_{n \in A} \frac{1}{n} = +\infty.$$

Então existem infinitas progressões aritméticas de comprimento arbitrário formada somente por elementos de A.

Esta conjectura está completamente em aberto: não se sabe nem se tais conjuntos devem conter progressões aritméticas de comprimento 3.

1.4 Prova do Teorema dos Números Primos

Nesta seção daremos um esboço da prova do teorema dos números primos. Veremos suas relações com a função de Von Mangoldt e com a função zeta de Riemann.

1.4.1 A função de Von Mangoldt

Em primeiro lugar, reformularemos o teorema numa linguagem integral e veremos suas relações com a função de Von Mangoldt.

Definição 1.4.1. *A função de Von Mangoldt* $\Lambda : \mathbb{Z} \to \mathbb{R}^+$ *é dada por* $\Lambda(n) = \log p$ *se* $n = p^r$*, onde* $r \geq 1$ *e* $\Lambda(n) = 0$ *caso contrário.*

Observe que o teorema da decomposição única em fatores primos pode ser expresso por:

$$\log n = \sum_{d|n} \Lambda(d). \tag{1.4.1}$$

Definição 1.4.2. *Dada* $f : X \to \mathbb{R}$ *e* $A \subset X$ *um conjunto finito, definimos a esperança de* f *com respeito à* A *como:*

$$\mathbb{E}(f(n)|n \in A) = \mathbb{E}(f|A) = \frac{1}{|A|} \sum_{n \in A} f(n).$$

Nesta linguagem o teorema dos números primos pode ser visto como uma estimativa para a esperança da função de von Mangoldt:

Teorema 1.4.1. *O Teorema dos Números Primos é equivalente à:*

$$\mathbb{E}(\Lambda|[1, N]) = 1 + o(1).$$

Demonstração. Pela definição da função de von Mangoldt temos que:

$$\begin{aligned} N\mathbb{E}(\Lambda|[1, N]) &= \sum_{p \leq N} [\frac{\log N}{\log p}] \log p \leq \log N \sum_{p \leq N} 1 \\ &= \log N \cdot (|\text{primos entre 1 e } N|). \end{aligned}$$

Isto dá uma das desigualdades desejadas (dividindo por N).

Por outro lado, se $1 < M < N$ então:

$$\begin{aligned} |\text{primos entre 1 e } N| &= |\text{primos entre 1 e } M| + \sum_{M < p \leq N} 1 \\ &\leq |\text{primos entre 1 e } M| + \sum_{M < p \leq N} \frac{\log p}{\log M} \\ &< M + \frac{1}{\log M} N\mathbb{E}(\Lambda|[1, N]). \end{aligned}$$

Agora se N é muito grande então temos que $1 < M = \frac{N}{\log^2 N} < N$. Substituindo na inequação acima obtemos que:

$$|\text{primos entre 1 e } N| < \frac{N}{\log^2 N} + \frac{N\mathbb{E}(\Lambda|[1,N])}{\log N - 2\log\log N}.$$

Portanto:

$$\frac{|\text{primos entre 1 e } n|}{N} < \mathbb{E}(\Lambda|[1,N])(\frac{1}{\log N - 2\log\log N}) + \frac{1}{\log^2 N}.$$

Isto conclui a demonstração porque $\frac{\log x}{\log x - 2\log\log x} \to 1$ quando $x \to \infty$. □

1.4.2 A função zeta de Riemann

Uma das funções mais famosas na Matemática é a função zeta de Riemann. Ela desempenha um papel fundamental na teoria dos números e também aparece em diversas outras áreas (p.ex., análise complexa, sistemas dinâmicos, etc.). Em particular, ela tem estreita relação com a distribuição dos números primos devido à fórmula de Euler. Nesta seção iremos estudar algumas propriedades desta função.

Definição 1.4.3. *A função zeta de Riemann é dada pela única extensão meromorfa da seguinte função analítica no domínio* $\{Re(s) > 1\}$*:*

$$\zeta(s) = \sum_{n\geq 1} \frac{1}{n^s}.$$

Proposição 1.4.1 (Fórmula de Euler)**.** *Se* $s > 1$ *é real, então a seguinte identidade de Euler é verdadeira:*

$$\zeta(s) = \prod_p \frac{1}{1 - p^{-s}}.$$

Para provar esta fórmula precisamos de falar de funções multiplicativas. Dizemos que uma função $f : \mathbb{Z} \to \mathbb{R}$ é multiplicativa se $f(mn) = f(m)f(n)$ quando $(m, n) = 1$, e ela é estritamente multiplicativa quando esta relação vale sem restrição.

Demonstração. Suponha que f é multiplicativa e limitada, então em $Re(s) > 1$ podemos escrever:

$$\sum \frac{f(n)}{n^s} = \prod_p (\sum_m f(p^m)p^{ms}).$$

De fato, em $Re(s) \geq \delta > 1$ temos que $\sum_{k\geq 1} \frac{f(k)}{k^s}$ converge uniformemente, por outro lado, seja $P(n) = \prod_{p\leq n}(\sum_{m\geq 0} f(p^m)p^{-ms})$, onde s está fixo. Ora, $P(n)$ é um produto finito de séries convergentes e podemos então escrevê-lo como $\sum_{m\in A_n} \frac{f(m)}{m^s}$, onde

$$A_n = \{r \in \mathbb{N};\ \text{os fatores primos de } r \text{ são menores ou iguais à } n\}.$$

Por exemplo se $n = 3$ temos que:

$$\begin{aligned} P(3) &= (\sum_m f(2^m)2^{-ms})(\sum_j f(3^j)3^{-js}) \\ &= \sum_{m,j} f(2^m)f(3^j)2^{-ms}3^{-js} \\ &= \sum_{m,j} f(2^m3^j)(2^m3^j)^{-s} = \sum_{m\in A_n} f(m)m^{-s}. \end{aligned}$$

Agora, obviamente $\{1,\ldots,n\} \subset A_n$. Logo $|P(n) - \sum_{k\geq 1} \frac{f(k)}{k^s}| \leq \sum_{k>n} \frac{f(k)}{k^s}$, donde o resultado segue pela convergência absoluta da série.

Além disso, se f é estritamente multiplicativa, temos que $f(p^m) = (f(p))^m$. Isto implica que $\sum_{m\geq 0} f(p^m)p^{-ms} = \sum_{m\geq 0}(f(p)p^{-s})^m = \frac{1}{1-f(p)p^{-s}}$, pela fórmula da série geométrica.

A fórmula de Euler segue então observando que a função $f \equiv 1$ é estritamente multiplicativa. □

Para nossos propósitos, estaremos interessados em conhecer regiões onde a função zeta não se anula. De fato, isso faz parte de um problema importante à respeito da função zeta conhecido como a *Hipótese de Riemann*: é sabido que os pares negativos $-2, -4, \ldots$ são zeros da função zeta, chamados de *zeros triviais*, e outros zeros conhecidos

são da forma $\frac{1}{2} + i\alpha$ onde α é um zero da função

$$\xi(t) = \frac{1}{2}s(s-1)\pi^{-s/2}\Gamma(\frac{s}{2})\zeta(s), \text{ onde } s = \frac{1}{2} + it.$$

A Hipótese de Riemann afirma que

Os zeros não triviais da função zeta de Riemann tem parte real igual à $\frac{1}{2}$.

Este é um problema muito difícil e tem posto a prova os esforços de muitos matemáticos famosos. Neste livro iremos encontrar uma região livre de zeros de forma elementar (usando análise complexa). A prova do seguinte fato será dada no apêndice deste capítulo:

Proposição 1.4.2. *$\zeta(s) \neq 0$ em $Re(s) \geq 1$.*

Para isso iremos estudar a analiticidade da função zeta de Riemann e provaremos o seguinte fato: *1 é o único polo da função zeta de Riemann em $Re(s) > 0$, ele é simples e tem resíduo 1. De fato, se $[x]$ denota a função maior inteiro menor do que x então na região $Re(s) > 0$ temos a expansão:*

$$\zeta(s) = \frac{1}{s-1} + 1 + s\int_1^\infty \frac{([x]-x)}{x^{1+s}}dx.$$

1.4.3 Prova Analítica

A principal ferramenta nesta demonstração é o seguinte teorema, cuja prova será dada no apêndice:

Teorema 1.4.2. *Seja $f : [1,\infty] \to \mathbb{R}$, tal que $f(x) = O(x)$, não descrescente e $f \in L^1_{loc}$. Dado s um parâmetro complexo, seja g a transformada de Mellin de f, isto é, $g(s) := s\int_1^\infty f(x)x^{-1-s}dx$. Então, em $Re(s) > 1$, g é uma função analítica. Além disso, se existe uma constante c tal que $g(s) - \frac{c}{s-1}$ tem continuação analítica em $\{Re(s) = 1\}$ então:*

$$\lim_{x\to\infty} \frac{f(x)}{x} = c.$$

Em vista do teorema 1.4.2 e das propriedades da função zeta, vistas acima, resta provar que $xE(\Lambda|[1,x]) = O(x)$. Podemos provar isto da seguinte maneira: observe que

$$\prod_{n<p\leq 2n} p \leq 2n(2n-1)\dots(n+1) = \binom{2n}{n} < 2^{2n}.$$

Assim, temos $\sum_{n<p\leq 2n} \log p < 2n\log 2$. Tomando $n = 2^{2k-1}$ (uma espécie de decomposição diádica) vemos que $\sum_{2^{k-1}<p\leq 2^k} \log p \leq 2^k \log 2$. Somando estas desigualdades temos que:

$$\sum_{p\leq 2^k} \log p \leq \sum_{1}^{k} 2^i \log 2 < 2^{k+1}\log 2.$$

Logo, tomando $2^{k-1} < N \leq 2^k$ segue que

$$\sum_{p\leq N} \log p \leq 2^{k+1}\log 2 = (4\log 2)2^{k-1} < 4N\log 2.$$

Portanto $\sum_{p^m\leq N} \log p < (4\log 2)N^{1/m}$. Agora, lembrando que $\mathbb{E}(\Lambda|[1,N]) = \frac{1}{N}(\sum_{p\leq N}\log p + \sum_{m\geq 2}\sum_{p^m\leq N}\log p)$ e definindo $\alpha^{-1} = [\frac{\log N}{\log 2}]$, obtemos:

$$\begin{aligned}\mathbb{E}(\Lambda|[1,N]) &\leq \frac{4\log 2}{N}(N + \sum_{m=2}^{1/\alpha} N^{1/2}) \\ &\leq 4\log 2(1+\frac{1}{\alpha\sqrt{N}}).\end{aligned}$$

Isto mostra a limitação desejada.

Estamos então nas hipótese do teorema 1.4.2, o qual diz que $\mathbb{E}(\Lambda|[1,N]) = 1+o(1)$. Mas isto é equivalente ao teorema dos números primos pelo teorema 1.4.1, o que conclui o argumento.

1.5 O Teorema de Van der Waerden

Nesta seção daremos duas provas do teorema de Van der Waerden:

Teorema 1.5.1 (Van der Waerden)**.** *Se colorirmos os inteiros positivos com um número k de cores, podemos achar progressões aritméticas de comprimento arbitrário formadas por somente uma cor.*

1.5.1 Prova Combinatória

Nesta seção, iremos provar o teorema de Van der Waerden através do método de colorir em Combinatória. Para não carregar muita notação, vamos denotar a progessão aritmética $a, a+r, \ldots, a+(k-1)r$ por $a+[0,k)r$, e vamos supor que temos m cores com as quais iremos colorir os números naturais de 1 até N.

Definição 1.5.1. *Seja $c : \{1, \ldots, N\} \to \{1, \ldots, m\}$ uma maneira de colorir. Dados $k \geq 1$, $d \geq 0$ e $a \in \{1 \ldots, N\}$, um ventilador de raio k, grau d com ponto base a é uma d-upla de progressões aritméticas $(a + [0,k)r_1, \ldots, a + [0,k)r_d)$ onde $r_1, \ldots, r_d > 0$. Para cada $1 \leq i \leq d$ as progressões $a + [1,k)r_i$ são chamadas de pás do ventilador. Dizemos que o ventilador é policromático se seu ponto base e suas pás são monocromáticas. Isto é, existem cores $c_0, c_1, \ldots, c_d$ distintas tais que $c(a) = c_0$ e $c(a + jr_i) = c_i$ para $j = 1, \ldots k$ e $i = 1, \ldots d$.*

Observe que pela distinção das cores, se temos m cores, é impossível termos um ventilador policromático com grau maior ou igual à m.

É claro que o teorema de van der Waerden segue do seguinte resultado:

Teorema 1.5.2. *Sejam $k, m \geq 1$. Então existe N tal que qualquer coloração com m cores de $\{1, \ldots, N\}$ contém uma progressão aritmética de comprimento k monocromática.*

Demonstração. A prova será feita em dois passos indutivos. Primeiro, faremos indução em k: observe que o caso $k = 1$ é trivial; tomemos $k \geq 2$ e vamos supor que o teorema é verdade para $k - 1$.

Em seguida faremos indução em d. Isto é, afirmamos que dado d, existe N tal que qualquer coloração com m cores de $\{1, \ldots, N\}$ contêm ou uma progressão aritmética monocromática ou um ventilador policromático de raio k e grau d. Note que o caso $d = 0$ é trivial e se provarmos que isso vale para $d = m$ então pela observação feita anteriormente, obtemos a progressão monocromática.

Vamos tomar $d \geq 1$ e supor que a afirmação vale para $d-1$. Seja $N = 4kN_1N_2$, onde N_1 e N_2 serão escolhidos depois e $A = \{1, \ldots, N\}$. Seja então $c : \{1, \ldots, N\} \to \{1, \ldots, m\}$ a coloração. Obviamente $\{bkN_1 + 1, \ldots, bkN_1 + N_1\}$ é um subconjunto de A com N_1 elementos para $b = 1, \ldots N_2$. Pela hipótese de indução em k e d se N_1 é muito grande, existe este conjunto possui uma progressão monocromática de comprimento k ou um ventilador policromático de raio k e grau $d-1$. Se para algum b encontramos a progressão monocromática, acabou. Portanto, vamos supor que para todo $b = 1, \ldots N_2$ sempre encontramos um ventilador policromático.

Logo, para cada $b = 1, \ldots, N_2$ encontramos $a(b), r_1(b), \ldots, r_{d-1}(b) \in \{1, \ldots, N_1\}$ e cores *distintas* $c_0(b), c_1(b), \ldots, c_{d-1}(b) \in \{1, \ldots m\}$ tais que $c(bkN_1 + a(b)) = c_0(b)$ e $c(bkN_1 + a(b) + jr_i(b)) = c_i(b)$ para $j = 1, \ldots, k-1$ e $i = 1, \ldots, d-1$. Chamaremos estas condições de primeira e segunda propriedades do ventilador gerado por b. Em particular, o mapa $b \to (a(b), r_1(b), \ldots, r_{d-1}(b), c_0(b), \ldots, c_{d-1}(b))$ é uma coloração com $m^d N_1^d$ cores do conjunto $\{1, \ldots, N_2\}$. Novamente pela hipótese de indução em k, se N_2 é muito grande, existe uma progressão monocromática $b + [0, k-1)s$ nesta nova coloração com alguma cor da forma $(a, r_1, \ldots, r_{d-1}, c_1, \ldots, c_{d-1})$. Revertendo a posição da progressão, podemos assumir que s é negativo, se for necessário.

A idéia agora é transformar uma progressão de ventiladores idênticos em um novo ventilador com um grau a mais, para completar o passo de indução. Seja então $b_0 = (b-s)kN_1 + a \in \{1, \ldots, N\}$ e considere o ventilador:

$$(b_0 + [0, k)skN_1, b_0 + [0, k)(skN_1 + r_1), \ldots, b_0 + [0, k)(skN_1 + r_{d-1}))$$

de raio k, grau d e ponto base b_0.

Vamos verificar que as pás são monocromática. Na primeira pá temos $c(b_0 + jskN_1) = c((b + (j-1)s)kN_1 + a)$ por substituição. Pela primeira propriedade do ventilador gerado por $b + (j-1)s$ segue que $c((b + (j-1)s)kN_1 + a) = c_0(b + (j-1)s) = c_0(b)$ (pois a progressão $b + [0, k-1)s$ é monocromática se $1 \leq j \leq k-1$). Da mesma forma, em uma pá arbitrária, usando a segunda propriedade do ventilador, temos que se $1 \leq j \leq k-1$ e $1 \leq t \leq d$ então:

$$c(b_0 + j(skN_1 + r_t)) = c((b + (j-1)s)kN_1 + a + jr_t) = c_t(b + (j-1)s) = c_t.$$

Se o ponto base b_0 tem a mesma cor de uma pá, então encontramos uma progressão monocromática de tamanho k, caso contrário o ponto base tem cor distinta de suas pás e portanto encontramos um ventilador policromático de raio k e grau d. Isto termina o passo indutivo e a prova do teorema. □

1.5.2 Prova via Sistemas Dinâmicos

Uma das grandes ferramentas em sistemas dinâmicos é a chamada dinâmica simbólica, a qual consiste em estudar uma transformação chamada *shift*. A seguir daremos a definição de shift e veremos como Furstenberg usou tal maquinária para dar uma prova do teorema de van der Waerden.

Seja $A = \{a_1, \dots, a_k\}$ um alfabeto finito. Considere todas as palavras infinitas compostas por letras deste alfabeto:

$$\Omega = \{(x_1, x_2, \dots, x_n, \dots) \; ; \; x_i \in A\}.$$

Este conjunto pode ser visto como um espaço métrico, através da seguinte distância: dados $x = (x_1, x_2, \dots)$ e $y = (y_1, y_2, \dots)$, defina

$$d(x, y) := \frac{1}{l} \text{ se } l \text{ é o menor inteiro tal que } x_l \neq y_l.$$

O shift é a transformação $T : \Omega \to \Omega$ definida por:

$$T(x_1, x_2, x_3, \dots) = (x_2, x_3, x_4, \dots).$$

É simples mostrar que o shift é uma aplicação contínua com respeito a métrica definida acima.

Com estes conceitos, Furstenberg usou o seguinte teorema de dinâmica topológica (cuja prova será dada no apêndice) para demonstrar o teorema de Van der Waerden.

Teorema 1.5.3 (Recorrência Múltipla Topológica - Furstenberg e Weiss)**.** *Seja $T : X \to X$ contínua e X um espaço métrico compacto. Para todo $k \in \mathbb{N}$ e $\varepsilon > 0$ existe $x \in X$ e $n \in \mathbb{N}$ tal que $d(T^{in}(x), x) < \varepsilon$ para todo $i = 1, \dots, k$. Mais ainda, dado $Z \subset X$ denso, podemos escolher $x \in Z$.*

Vejamos como podemos provar o teorema de Van der Waerden a partir deste resultado. Seja $A = \{c_1, \dots, c_k\}$ o conjunto de cores e $z = (z_1, z_2, z_3, \dots)$ uma maneira de colorir $\mathbb{N}$ onde $z_i \in A$ indica a cor do número i. Consideremos então $z \in A^{\mathbb{N}}$ e $T : A^{\mathbb{N}} \to A^{\mathbb{N}}$ o shift. Lembrando a definição da distância, temos que, para $x, y \in A^{\mathbb{N}}$ e $m, l \in \mathbb{N}$, vale $d(T^m(x), T^l(y)) < 1$ se e só se $x_{m+1} = y_{l+1}$.

Em particular, se $z \in A^{\mathbb{N}}$ então a progressão aritmética $m, m+n, \dots, m+kn$ é monocromática se $z_m = z_{m+n} = \cdots = z_{m+kn}$, ou seja, se:

$$\begin{aligned} d(T^{m-1}(z), T^{m-1+in}(z)) &= d(T^{m-1}(z), T^{in}(T^{m-1}(z))) \\ &< 1, \text{ para } i = 1, \dots k. \end{aligned}$$

Tomando $X = \overline{\{T^m(z)\}_{m=0}^{\infty}}$, temos que X é compacto, T é contínua em X e o conjunto $Z = \{T^m(z)\}_{m=0}^{\infty}$ é denso em X. O teorema de Van der Waerden segue então do teorema 1.5.3.

1.6 O Teorema de Furstenberg e suas aplicações no teorema de Szemerédi

Nesta seção daremos uma prova do teorema de Szemerédi baseada em elementos de teoria ergódica (mais ou menos inspirados pela "prova dinâmica" do teorema de van der Waerden). Primeiramente faremos uma introdução aos conceitos básicos de teoria ergódica, em seguida enunciaremos o teorema de recorrência múltipla ergódica de Furstenberg e, como corolário, obteremos o teorema de Szemerédi.

1.6.1 Breve Introdução à Teoria Ergódica

A teoria ergódica estuda iterações de uma transformação $T : X \to X$, onde X é um espaço de medida, do ponto de vista de uma medida μ *invariante* pela transformação T (i.e., para todo subconjunto mensurável A temos que $\mu(A) = \mu(T^{-1}(A))$).

A presença da medida invariante dá muita informação estatística sobre a estrutura de órbitas da transformação T, isto é, dos conjuntos $\{T^n(x)\}_{n=0}^{\infty}$, para quase todo $x \in X$ com respeito a medida μ. Por exemplo, o teorema de Poincaré diz que "se $T : X \to X$ é μ-invariante

e $\mu(A) > 0$ então para μ-qtp x em A temos que existe um $n(x) \geq 1$ tal que $T^{n(x)}(x) \in A$". Portanto existe um N tal que:

$$\mu(A \cap T^{-N}(A)) > 0.$$

Em particular, vemos que, por menor que seja um conjunto contendo um ponto x, se ele tem medida positiva então existem muitas órbitas que começam dentro desse conjunto e voltam infinitas vezes para este conjunto. Se o espaço de medida for topológico, então podemos reformular o teorema de Poincaré da seguinte maneira:

Sejam $T : X \to X$ um espaço de medida e métrico (com respeito à uma métrica d) e μ uma medida invariante por T. Então quase todo ponto com respeito à μ é recorrente, isto é para quase todo ponto x existe uma sequência $n_k \to \infty$ de naturais tais que $d(T^{n_k}(x), x)) \to 0$ quando $k \to \infty$.

Uma pergunta natural é se existem sempre medidas invariantes para alguma tranformação T dada. Quando o espaço X é compacto e a transformação é contínua, a resposta é sim. A idéia da prova deste fato é muito simples: tomemos uma medida qualquer arbitrária e vejamos como essa medida muda pela ação da transformação, ou melhor pela ação de iterados da transformação. Faça uma média dessas medidas até o iterado N; conforme N cresce, essa nova medida tende a ficar menos sensível a ação de T. O ponto é tomar estudar o limite quando $N \to \infty$ e torcer para que uma medida limite exista; de acordo com nosso argumento (informal) tal limite será invariante por T.

Façamos agora a construção com mais detalhes. Como vimos no parágrafo anterior, iremos tomar um certo limite de medidas, de maneira que precisamos de um conceito de convergência de medidas. Como o espaço de medidas de Radon é o dual do espaço de funções contínuas é natural usarmos a topologia fraca, uma vez que pela Análise Funcional teremos de graça resultados de compacidade (ajudando na questão da existência de um "limite").

Definição 1.6.1. *Dizemos que uma sequência de medidas μ_k em X converge fracamente para μ se para toda função contínua $f : X \to \mathbb{R}$ vale:*

$$\int_X f d\mu_k \to \int_X f d\mu.$$

Como esta topologia é a topologia fraca, temos pelo teorema de Banach-Alaoglu que:

O espaço de probabilidades em X (isto é, o conjunto de medidas μ tais que $\mu(X) = 1$) é compacto com respeito a convergência fraca.

Voltemos agora para a questão da existência de medidas invariantes. Seja η uma probabilidade *qualquer*. A ação dos iterados de T na medida η será dada pelo *push-forward*, isto é, $((T^n)^*\eta)(A) := \eta(T^{-n}(A))$ para todo conjunto mensurável A. Uma observação importante é que a propriedade de uma medida η ser invariante pode ser traduzida na equação $T^*\eta = \eta$.

Vamos considerar a seguinte sequência de probalidades:

$$\mu_k = \frac{1}{k}\sum_{i=0}^{k-1}(T^i)^*\eta.$$

Em suma, estamos tomando médias temporais das medidas obtidas por push-forward. Por compacidade, vemos que existe uma subsequência μ_{n_k} que converge fracamente para alguma probabilidade μ. Afirmamos que μ é invariante. De fato, temos as seguintes igualdades que serão explicadas logo em seguida:

$$\begin{aligned}
T^*\mu &= T^*(\lim \mu_{n_k}) \\
&= \lim(T^*(\mu_{n_k})) \\
&= \lim(\frac{1}{n_k}\sum_{i=0}^{n_k-1}(T^{i+1})^*(\eta)) \\
&= \lim(\frac{1}{n_k}(\sum_{i=0}^{n_k}(T^i)^*(\eta) - \eta + (T^{n_k})^*\eta)) \\
&= \lim\frac{1}{n_k}\sum_{i=0}^{n_k}(T^i)^*(\eta) \\
&= \mu.
\end{aligned}$$

Na segunda igualdade usamos o fato que o operador T^* é contínuo na topologia fraca, pois T é contínua. Com efeito, suponha que

$\mu_k \to \mu$ fracamente e fixe $f : X \to \mathbb{R}$ contínua. Então temos que $f \circ T$ também é contínua e portanto:

$$\int_X f d(T^*\mu_k) = \int_X f \circ T d\mu_k \to \int_X f \circ T d\mu = \int_X f d(T^*\mu).$$

Para a quinta igualdade, observamos que, para toda $f : X \to \mathbb{R}$ contínua, temos, por compacidade de X:

$$\frac{1}{n_k}\int_X f d\mu \to 0 \text{ e } \frac{1}{n_k}\int_X f d((T^{n_k})^*\mu) = \frac{1}{n_k}\int_X f \circ T^{n_k} d\mu \to 0.$$

Logo as duas últimas parcelas convergem à zero fracamente.

Um exemplo concreto interessante para nossos propósitos futuros é $X = \{0,1\}^{\mathbb{N}}$ e T o shift. Considerando a medida de Dirac de um ponto $x \in X$, ou seja, $\delta_x(A) = 0$ se $x \notin A$ e $\mu(A) = 1$ se $x \in A$, então a sequência $\mu_k = \frac{1}{k}\sum_{j=0}^{k-1} \delta_{T^j(x)}$ possui um ponto de acumulação na topologia fraca e este ponto é uma medida invariante pelo shift.

1.6.2 O teorema de Furstenberg

Uma pergunta natural a respeito do teorema de Poincaré é se dado o conjunto A com medida positiva existe uma certa estrutura no conjunto de iterados que retornam à A; mais precisamente, sabemos que o conjunto é infinito, mas será que existe uma estrutura aritmética neste conjunto? Esta pergunta foi resolvida por Furstenberg e sua resposta é conhecida como o teorema de Recorrência Múltipla Ergódica de Furstenberg.

Teorema 1.6.1 (Recorrência Múltipla Ergódica de Furstenberg)**.** *Seja $T : X \to X$ μ-invariante, $k \geq 3$ e $\mu(A) > 0$ então existe N tal que:*

$$\mu(A \cap T^{-N}(A) \cap \cdots \cap T^{-(k-1)N}(A)) > 0.$$

Este teorema é o coração da prova do teorema de Szemerédi atráves de métodos da Teoria Ergódica. Para indicar ao leitor a natureza deste resultado, veremos agora a prova do teorema de Furstenberg em certos casos particulares importantes.

O primeiro exemplo é tomar um sistema de Bernoulli. Novamente, seja A um conjunto finito com r elementos, $X = A^{\mathbb{N}}$ e T o shift neste

espaço. Sejam $p_1, \ldots, p_r$ números não-negativos tais que $\sum p_i = 1$. Isto dá uma probabilidade em A e tomando a medida produto temos uma probabilidade em X.

A σ-álgebra produto é gerada pelos cilindros com um número finito n de coordenadas, isto é conjuntos da forma $C = \{w \in Z; w_{i_1} = j_1, \ldots w_{j_n} = j_n\}$ e a medida de Bernoulli é dada por $\mu(C) = p_{j_1} \ldots p_{j_n}$ sendo depois estendida para a σ-álgebra gerada. É simples mostrar que esta medida é invariante pelo shift (de fato basta mostrar que $\mu(B) = \mu(T^{-1}(B))$ apenas quando B é um cilindro).

Da mesma maneira, como os cilindros geram a σ-álgebra, basta provar o teorema de Furstenberg para tais conjuntos. Sejam então $C_0, C_1, \ldots, C_k$ cilindros e observe que se n é muito grande então as coordenadas que definem os cilindros $T^{-nl}(C_l)$ são todas disjuntas. Portanto, temos:

$$\mu(C_0 \cap T^{-n}(C_1) \cap \cdots \cap T^{-kn}(C_k)) = \mu(C_0)\mu(C_1) \ldots \mu(C_k)) > 0.$$

Isto prova o teorema neste exemplo.

Outro exemplo seria um sistema periódico, isto é, uma dinâmica tal que $T^p = T$ para algum p. Neste caso, o resultado é totalmente trivial; uma dinâmica (menos trivial) nesta linha de raciocínio é o seguinte exemplo quase-periódico: $X = S^1 = \mathbb{R}/\mathbb{Z}$, μ a medida de Lebesgue no círculo e $T(x) = x + \alpha (\text{mod } 1)$ para algum α.

Dado A um conjunto mensurável tal que $\mu(A) > 0$, note que a função $\int 1_A(x + y)d\mu(x)$ é contínua em y. Logo, para todo $\varepsilon > 0$ existe δ tal que se $|y| < \delta$ então $\mu(A \cap (A - y)) > \mu(A) - \varepsilon$. Portanto:

$$\mu(A \cap (A - y) \cap (A - 2y) \cap \cdots \cap (A - ky)) > \mu(A) - (k + 1)\varepsilon.$$

Escolhendo $\varepsilon < \frac{\mu(A)}{k+1}$, tomando o δ correspondente e definindo o conjunto $D_\delta = \{n \geq 1; n\alpha \in (-\delta, \delta)(mod\ 1)\}$, então, se $n \in D_\delta$ temos que:

$$\mu(A \cap T^{-n}(A) \cap \cdots \cap T^{-nk}(A)) > \mu(A) - (k + 1)\varepsilon > 0.$$

Notemos que o primeiro exemplo é um caso particular de sistemas *fracamente misturadores* ("weak-mixing"), isto é sistemas que satisfazem a seguinte igualdade, para todo A e B subconjuntos men-

suráveis:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} (\mu(A \cap T^{-n}B) - \mu(A)\mu(B))^2 = 0.$$

Adaptando a idéia do caso Bernoulli, prova-se que o teorema de Furstenberg vale para sistemas fracamente misturadores (que podemos chamar também de caso "pseudo-aleatório") .

O segundo exemplo é um caso particular de sistemas compactos, isto é sistemas tais que, para toda função $f \in L^2(\mu)$, o fecho do conjunto $\{f, Tf, T^2f, \dots, T^nf, \dots\}$ é compacto. Adaptando a idéia usada para rotações, prova-se o teorema de Furstenberg para sistemas compactos (que podemos chamar também de caso "estruturado").

No contexto geral, o teorema de Furstenberg segue então de uma decomposição em vários níveis do sistema em partes fracamente misturadoras e compacta, que não tem muita correlação entre si ao longo de uma torre de extensões. Cabe ressaltar que a existência desta torre de extensões é um fato altamente não-trivial (conhecido como teorema de estrutura) e foge ao escopo deste livro. Entretanto, uma versão desta idéia num contexto finitário será exposta no capítulo 2 do livro.

1.6.3 Prova do teorema de Szemerédi

Uma vez com o teorema de recorrência múltipla podemos dar uma prova rápida do teorema de Szemerédi usando o shift.

Sejam $X = \{0,1\}^{\mathbb{N}}$ e $T : X \to X$ o shift. Tome $(x_n) = 1_A(n)$, onde $1_A(x)$ é a função característica de A, e considere $\mu_k = \frac{1}{k}\sum_{j=0}^{k-1} \delta_{T^j(x)}$. Então, como vimos anteriormente, a menos de passar a uma subsequência, podemos supor que $\mu = \lim \mu_k$ é uma medida invariante para T.

Defina $Y = \{(y_n); y_1 = 1\}$. Temos $\mu(Y) = \lim \mu_k(Y) = \lim \frac{1}{k}|A \cap [1,k]| > 0$ por hipótese. Logo, pelo teorema de Furstenberg, segue que existe um N tal que $\mu(Y \cap T^{-N}(Y) \cap \cdots \cap T^{-(k-1)N}(Y)) > 0$. Em particular existem (infinitos) $z \in Y \cap T^{-N}(Y) \cap \cdots \cap T^{-(k-1)N}(Y)$. Isto é, existe x tal que $x, x+N, \dots, x+(k-1)N \in A$. Isto prova o teorema de Szemerédi.

1.7 O Teorema de Szemerédi quantitativo

Nesta seção veremos algumas reformulações do teorema 1.3.1 de Szemerédi (muito úteis para os nossos propósitos futuros).

Começaremos por observar que o teorema de Szemerédi é *equivalente* a seguinte afirmação:

Para todo $k \geq 1$ e $0 < \delta \leq 1$, existe um inteiro $N_{SZ}(k, \delta) \geq 1$ tal que, para qualquer $N \geq N_{SZ}$, todo conjunto $A \subset \{1, \ldots, N\}$ com cardinalidade $|A| \geq \delta N$ contém pelo menos uma progressão aritmética de comprimento k.

Logicamente, a afirmação acima implica diretamente o teorema de Szemerédi.

Na outra direção, mostraremos agora que, se a afirmação é falsa para um certo par (k, δ) então existe um conjunto $Y \subset \mathbb{N}^*$ com $|Y \cap \{1, 2, \ldots, n\}| \geq \delta r$, $\forall\, r \in \mathbb{N}^*$ tal que Y não contém nenhuma progressão aritmética de comprimento k.

Para isso, provaremos inicialmente que, se não existe $N_{SZ}(k, \delta)$, então, para cada $n \in \mathbb{N}^*$, existe um conjunto $X_n \subset \{1, 2, \ldots, n\}$ com $|X_n \cap \{1, 2, \ldots, k\}| \geq \delta k$ para $1 \leq k \leq n$ tal que X_n não contém nenhuma progressão aritmética de tamanho k. Com efeito, seja $\varepsilon_n = \max_{1 \leq k \leq n} ((\lceil \delta k \rceil - 1)/k) < \delta$. Afirmamos que, se N é suficientemente grande e $A \subset \{1, 2, \ldots, N\}$ tem pelo menos δN elementos, então existe $m \leq N - n$ tal que $|A \cap \{m + 1, \ldots, m + k\}| \geq \delta k$, para $1 \leq k \leq n$. De fato, se não for o caso, existem $s \in \mathbb{N}^*$, $k_1, k_2, \ldots, k_s \in \{1, 2, \ldots, n\}$ tais que $N \geq k_1 + k_2 + \cdots + k_s > N - n$ e, para $1 \leq r \leq s$, $|A \cap \big(\sum_{j<r} k_j, \sum_{j \leq r} k_j\big]| < \delta k_r$,, donde $\frac{1}{k_r} |A \cap \big(\sum_{j<r} k_j, \sum_{j \leq r} k_j\big]| \leq \varepsilon_n < \delta$, e logo $\delta N \leq |A| \leq n + \varepsilon_n \cdot N$, o que é absurdo para $N > \frac{n}{\delta - \varepsilon_n}$. Logo, basta tomar um conjunto $A \subset \{1, 2, \ldots, N\}$ com pelo menos δN elementos que não contém nenhuma progressão aritmética de tamanho k para concluir a existência de X_n.

Agora, para cada $r \in \mathbb{N}^*$, seja $\pi_r \colon 2^{\mathbb{N}} \to 2^{\{1,2,\ldots,r\}}$ dada por $\pi_r(A) = A \cap \{1, 2, \ldots, r\}$. Construímos indutivamente conjuntos $Y_1, Y_2, Y_3, \ldots$ com $Y_r \subset \{1,2,\ldots,r\}$ para cada $r \in \mathbb{N}^*$ tais que $Y_{r+1} \cap \{1, 2, \ldots, r\} = Y_r$, $\forall\, r \in \mathbb{N}^*$ da seguinte forma: $Y_1 := \{1\} \subset X_n$, para

todo $n \in \mathbb{N}^*$. Dado Y_r, $r \in \mathbb{N}^*$ tal que $Y_r = \pi_r(X_n)$ para infinitos $n \in \mathbb{N}$, existe $Y_{r+1} \subset \{1, 2, \ldots, r+1\}$ com $Y_{r+1} \cap \{1, 2, \ldots, r\} = Y_r$ tal que $Y_{r+1} = \pi_{r+1}(X_n)$ para infinitos $n \in \mathbb{N}$ (de fato, se $\pi_r(X_n) = Y_r$, existem apenas duas possibilidades para $\pi_{r+1}(X_n)$). Agora, é fácil ver que $Y = \bigcup_{n \in \mathbb{N}^*} Y_n$ satisfaz $\pi_r(Y) = Y_r$, $\forall\, r \in \mathbb{N}^*$, donde $\pi_r(Y) = \pi_r(X_n)$ para infinitos $n \in \mathbb{N}$. Em particular, $|Y \cap \{1, 2, \ldots, r\}| \geq \delta r$, $\forall\, r \in \mathbb{N}^*$ e Y não contém nenhuma progressão aritmética de tamanho k.

Em seguida, vamos introduzir uma linguagem mais analítica e finitária para obter uma outra reformulação do teorema 1.3.1. Com este intuito, relembremos a seguinte definição:

Definição 1.7.1. *Seja $f : A \to \mathbb{C}$ onde A é um conjunto finito. Então:*

$$\mathbb{E}(f) = \mathbb{E}(f(n); n \in A) = \frac{1}{|A|} \sum_{n \in A} f(n).$$

Dada $f : (\mathbb{Z}/N\mathbb{Z}) \to \mathbb{R}$ uma função, podemos definir $T^n f : (\mathbb{Z}/N\mathbb{Z}) \to \mathbb{R}$ os shifts da função por números naturais $n \in \mathbb{Z}/N\mathbb{Z}$ (ou $n \in \mathbb{Z}$) através da fórmula: $T^n f(x) := f(x + n)$. Diremos também que uma função $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ é limitada se $\|f\|_{L^\infty} \leq 1$.

Com esta notação, podemos reformular o teorema de Szemerédi do seguinte jeito:

Teorema 1.7.1 (Szemerédi – versão quantitativa)**.** *Para todo $k \geq 1$ inteiro e $0 < \delta \leq 1$ real, existem $N_0(k, \delta)$ inteiro e $c(k, \delta) > 0$ real tais que, para todo $N \geq N_0(k, \delta)$ um número primo grande, qualquer $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}^+$ uma função limitada com $\mathbb{E}(f|\mathbb{Z}/N\mathbb{Z}) \geq \delta$ verifica*

$$\mathbb{E}(\prod_{j=0}^{k-1} T^{jr} f(x) | x, r \in \mathbb{Z}/N\mathbb{Z}) \geq c(k, \delta).$$

Observação 1.7.1. *Com relação a versão quantitativa do teorema de Szemerédi enunciada acima, iremos construir (no apêndice deste capítulo) alguns exemplos devidos a F. Behrend de subconjuntos S do intervalo $[1, N]$ tais que $|S| \geq N^{1 - \frac{2\sqrt{2 \log 2} + \varepsilon}{\sqrt{\log N}}}$ e S não contém progressões aritméticas de tamanho 3. Mais ainda, modificando o*

esquema do argumento de Behrend, veremos que, acerca do comportamento de $c(k,\delta)$ acima em termos de δ, não podemos esperar em geral que $c(k,\delta)$ tenha comportamento polinomial em δ (i.e., $c(k,\delta) \geq \delta^{C_k}$ para algum $C_k > 0$): com efeito, provaremos que $c(3,\delta) \leq \delta^{c\log(1/\delta)}$.

Observe que o enunciado do teorema 1.7.1 fornece (*a príncipio*) uma conclusão muito mais poderosa que o teorema de Szemerédi usual. Com efeito, enquanto o teorema de Szemerédi permite concluir apenas a existência de *uma* k-PA, a versão quantitativa permite inferir a existência de $c(k,\delta)N^2$ k-PAs (ao menos). Entretanto, apesar do teorema quantitativo de Szeméredi aparentar ter um enunciado mais forte, afirmamos que os teoremas 1.3.1 e 1.7.1 são *equivalentes*.

Iniciaremos vendo porque o teorema de Szemerédi segue da versão quantitativa: fixe k, δ e tome N um primo bem grande. Vamos supor que $A \subset \{1,\ldots,N\}$ tem cardinalidade $|A| \geq \delta N$, pois A tem densidade positiva. Seja N' um primo entre kN e $2kN$ (cuja existência é assegurada pelo postulado de Bertrand). Vamos considerar $\{1,\ldots,N\}$ como subconjunto de $\mathbb{Z}/N'\mathbb{Z}$ e A' o conjunto respectivo de $\mathbb{Z}/N'\mathbb{Z}$.

Ora, nossas escolhas implicam $E(1_{A'}|\mathbb{Z}/N\mathbb{Z}) \geq \delta/2k$. Pelo teorema de Szemerédi quantitativo segue que:

$$E(\prod_{j=0}^{k-1} T^{jr}1_{A'}(x)|x,r \in \mathbb{Z}/N'\mathbb{Z}) \geq c(k,\delta/2k).$$

Reescrevendo, temos que:

$$|\{(x,r) \in (\mathbb{Z}/N'\mathbb{Z})^2; x, x+r, \ldots, x+(k-1)r \in A'\}| \geq c(k,\delta/2k)(N')^2.$$

Como $N' \geq kN$ e $A' \subset \{1,\ldots,N\}$ temos que $1 \leq x \leq N$ e $-N \leq r \leq N$. Observe as progressões com $r = 0$ contribuem apenas com no máximo N elementos. Removendo estas progressões e tomando N grande, o lado direito da estimativa ainda é positivo e portanto A contém a progressão $x, x+r, \ldots, x+(k-1)r$. Vamos ver agora que, de fato, a versão quantitativa do teorema de Szemerédi é equivalente à sua versão original. Já sabemos que a versão original é equivalente à versão finita, i.e., à existência de $N_{SZ}(k,\delta)$, $\forall\, k \in \mathbb{N}^*$, $\delta > 0$. Logo, para concluir a equivalência entres as duas formulações do teorema de Szemerédi, basta mostrar a seguinte

Proposição 1.7.1. *Suponha que existe* $N_{SZ}\big(k, \frac{\delta}{2}\big)$. *Então existem* $N_0 \in \mathbb{N}$ *e* $\alpha(k,\delta) > 0$ *tais que, se* $N \geq N_0$ *para todo* $A \subset \{1, 2, \ldots, N\}$ *com* $|A| \geq \delta N$, *existem pelo menos* $\alpha(k,\delta)\, N^2$ *progressões aritméticas de comprimento* k *contidas em* A.

Demonstração. Seja $m_0 = N_{SZ}(k, \delta/2)$. Então, para todo $m \geq m_0$, todo subconjunto de $\{1, 2, \ldots, m\}$ com pelo menos $\delta m/2$ elementos contém uma progressão aritmética de comprimento k. Seja N grande. Para cada r com $1 \leq r \leq \lfloor N/m_0 \rfloor$, dividimos $\{1, 2, \ldots, N\}$ em r progressões aritméticas de razão r, do tipo $\{1 \leq n \leq N \mid n \equiv a (\mathrm{mod}\ r)\}$, para cada a com $0 \leq a \leq r-1$. Cada uma dessas PA's tem pelo menos $\lfloor N/r \rfloor$ elementos, e portanto pode ser decomposta como a união de $\lfloor \lfloor N/r \rfloor / m_0 \rfloor$ progressões aritméticas disjuntas de razões iguais a r, comprimentos $\geq m_0$ (e quase iguais) e portanto diâmetros entre $r(m_0-1)$ e $r(2m_0-1)$. Se $A \subset \{1, 2, \ldots, N\}$ satisfaz $|A| \geq \delta N$, para cada r, $\#\{0 \leq a \leq r-1 \mid \#A \cap \{1 \leq n \leq N \mid n \equiv a (\mathrm{mod}\ r)\} \geq \frac{3\delta}{4} \lfloor N/r \rfloor\} \geq \frac{\delta r}{4-3\delta}$ (pois $t < \frac{\delta}{4-3\delta} \Rightarrow t + \frac{3\delta}{4}(1-t) < \delta$), e (como $t < \frac{\delta}{4-2\delta} \Rightarrow t + \frac{\delta}{2}(1-t) < \frac{3\delta}{4}$) se $\#A \cap \{1 \leq n \leq N \mid A \equiv a (\mathrm{mod}\ r)\} \geq \frac{3\delta}{4} \lfloor N/r \rfloor$, pelo menos $\frac{\delta}{4-2\delta} \cdot \lfloor \lfloor N/r \rfloor / m_0 \rfloor$ das progressões de comprimento $\geq m_0$ que criamos têm interseção com A com proporção relativa pelo menos $\delta/2$, e logo contêm uma k-PA. Isto fornece pelo menos $\sum_{r=1}^{\lfloor N/m_0 \rfloor} \frac{\delta r}{4-3\delta} \cdot \frac{\delta}{4-2\delta} \lfloor \lfloor N/r \rfloor / m_0 \rfloor > \beta(\delta, m_0) N^2$ k-PA's contidas em A para N grande, onde $\beta(\delta, m_0) = \delta^2/64m_0^2$. Cada uma dessas PA's pode estar sendo contada algumas vezes, para diferentes escolhas de r, mas se d é seu diâmetro, r deve ser um divisor de d entre $\frac{d}{2m_0-1}$ e $\frac{d}{k-1}$, i.e., $r = \frac{d}{r'}$, onde $k-1 \leq r' \leq 2m_0-1$. Temos assim no máximo $2m_0 - k + 1$ possibilidades para r', e logo para r, i.e., cada PA é contada no máximo $2m_0 - k + 1$ vezes. Assim, A contém pelo menos $\alpha(k,\delta)N^2$ $k-$PA's, onde $\alpha(k,\delta) = \dfrac{\delta^2}{64m_0^2(2m_0-k+1)}$. □

Observação 1.7.2. *A diferença da prova do Teorema de Szemerédi quantitativo para as provas anteriores é que, devida a natureza finitária dos argumentos, podemos obter cotas explícitas para o número N_{SZ}. As outras provas, como são de caráter "infinito" (usam de certa maneira o Axioma da Escolha) somente mostram a existência de tal número e não dizem nada sobre a ordem de grandeza do mesmo. A estratégia da prova deste teorema foi usada no resultado de Green-Tao e será estudada no capítulo 2. Recomendamos também a leitura de [11].*

1.8 Outros resultados

Nesta seção indicaremos, sem provas, outras maneiras de enunciar algumas das conjecturas citadas acima. Em seguida apresentaremos alguns resultados posteriores ao teorema de Green-Tao. Finalmente, faremos alguns comentários sobre a natureza do número $N_0(k, \delta)$.

1.8.1 A função de Von Mangoldt e Reformulações de algumas conjecturas

Vimos nas seções anteriores, que o teorema dos números primos pode ser enunciado em termos da função de Von Mangold como:

$$\frac{1}{N}\sum_{n=1}^{N}\Lambda(n) = 1 + o(1).$$

Na verdade, melhorar as cotas para a esperança da função de von Mangoldt[1] implica em diversas conjecturas. Vamos listar algumas delas sem provas:

- A Hipótese de Riemann é equivalente à seguinte afirmação:

$$\mathbb{E}(\Lambda(n)|[1, N]) = 1 + O(N^{-1/2}\log^2 N).$$

- A conjectura dos primos gêmeos seguiria da seguinte afirmação:

$$\liminf_{N\to\infty} \mathbb{E}(\Lambda(n)\Lambda(n+2) : 1 \leq n \leq N) > 0.$$

[1] Isto é, explicitar a velocidade de convergência para zero do termo $o(1)$.

- A conjectura de Goldbach é equivalente à:

$$\mathbb{E}(\Lambda(n_1)\Lambda(n_2)|n_1, n_2 \in [1, N] \text{ e } n_1 + n_2 = N) > 0 \ \forall \ N \text{ par.}$$

- E, finalmente, a conjectura fraca de Goldbach é equivalente à:

$$\mathbb{E}(\Lambda(n_1)\Lambda(n_2)\Lambda(n_3)|n_1, n_2, n_3 \in [1, N] \text{ e } n_1+n_2+n_3 = N) > 0 \ \forall \ N \text{ ímpar.}$$

1.8.2 Constelações de Primos e Progressões Polinomiais

Um outro conjunto onde a noção de primalidade existe são os inteiros Gaussianos $\mathbb{Z}[i] := \{a + bi; a, b \in \mathbb{Z}\}$. Neste caso, p é um primo Gaussiano se ele só é divisível por $\pm 1, \pm i, \pm p$ e $\pm ip$.

Uma *forma* em $\mathbb{Z}[i]$ é um conjunto finito $(v_j)_{j\in J} \in (\mathbb{Z}[i])^J$ de inteiros Gaussianos distintos. Uma *constelação* em $\mathbb{Z}[i]$ com esta forma é qualquer J-upla $(a + rv_j)_{j\in J} \in (\mathbb{Z}[i])^J$ (onde $a \in \mathbb{Z}[i]$ e $r \in \mathbb{Z}[i]$) de inteiros Gaussianos distintos.

A noção de constelação estende a noção de progressões aritméticas para inteiros Gaussianos. A existência de muitas constelações formadas por primos Gaussianos foi demonstrada por Tao [12]:

Seja $(v_j)_{j\in J}$ uma forma qualquer de inteiros Gaussianos. Então os primos Gaussianos contêm inifinitas constelações com esta forma.

Por outro lado, uma maneira alternativa de generalizar o conceito de progressão aritmética é: como uma progressão aritmética toma a forma $x + P_1(m), \ldots, x + P_k(m)$ onde $P_i(m) = (i-1)m$, podemos estender esta definição permitindo que $P_i \in \mathbb{Z}[m]$ sejam polinômios com valores inteiros tais que $P_i(0) = 0$ (para $i = 1 \ldots k$). Estas progressões generalizadas são ditas *progressões polinomiais*.

A existência de infinitas progressões polinomiais formadas por primos foi demonstrada recentemente por Tao e Ziegler [13]:

Sejam $P_1, \ldots P_k$ polinômios como acima; dado $\varepsilon > 0$ existem infinitos inteiros x e m tais que $1 \leq m \leq x^{\varepsilon}$ e $x + P_i(m)$ são primos para $i = 1 \ldots k$.

1.8.3 Buracos no conjunto dos números primos

De certa forma, todos os teoremas e resultados que apresentamos aqui tem em comum a busca de certos padrões no conjunto dos números primos. Neste sentido, uma pergunta natural é quão esparsos são os primos.

Seja então p_n o n-ésimo primo, de modo que o tamanho do n-ésimo buraco do conjunto dos primos é $p_{n+1} - p_n$. O teorema dos números primos diz que a média do tamanho destes buracos é essencialmente $\log p_n$. Vamos definir Δ como o menor número tal que existem infinitos buracos de tamanho menor que $(\Delta + \varepsilon)$ vezes a média dos tamanhos. Isto é:

$$\Delta = \liminf_{n\to\infty} \left(\frac{p_{n+1} - p_n}{\log p_n} \right).$$

Conjecturava-se que $\Delta = 0$ e isto foi provado em trabalhos recentes: primeiro Goldston e Yıldırım [6][2] mostraram que $\Delta < \frac{1}{4}$ e, em seguida, Goldston, Motohachi, Pintz e Yıldırım [4] demonstraram a conjectura. Nestes trabalhos, eles propõe um método para mostrar a existência de números primos grandes muito próximos[3].

Observação 1.8.1. *Somente para ressaltar a dificuldade da conjectura dos primos gêmeos, observe que a conjectura dos primos gêmeos é uma afirmação muito mais forte do que o resultado $\Delta = 0$ de Goldston, Motohachi, Pintz e Yıldırım (o qual não é nada simples de se provar!).*

1.8.4 O tamanho do número $N_0(k, \delta)$

Sobre a magnitude do número $N_0(k, \delta)$ no teorema de Szemerédi quantitativo, temos os seguintes resultados:

- T. Gowers provou que $N_0(k, \delta) \leq 2^{2^{\delta^{-c_k}}}$, onde $c_k = 2^{2^{k+9}}$;
- R. Rankin provou que $N_0(k, \delta) \geq \exp(C(\log \frac{1}{\delta})^{1+\lfloor \log_2(k-1) \rfloor})$;

[2] Os pingos nos i´s não existem no nome de Yıldırım!

[3] De fato, eles mostram que estes primos estão realmente bem próximos assumindo uma conjectura de Elliot-Halberstam.

- J. Bourgain provou que $N_0(3,\delta) \leq 2^{C\delta^{-2}\log(1/\delta)}$;
- Espera-se que $N_0(k,\delta) \leq 2^{c_k\delta^{-1}}$, mas isto é um problema em aberto (relacionado a conjectura de Erdös-Turán).

1.9 Apêndice ao Capítulo 1

1.9.1 Prova do Teorema de Dirichlet no caso $a = 1$ e b qualquer

Nesta seção daremos uma prova deste caso particular usando polinômios ciclotômicos.

Sejam $\zeta_n = \cos(\frac{2\pi}{n}) + i\sin(\frac{2\pi}{n})$. Então, temos que $\zeta_n^k \neq 1$ para todo $k = 1, \ldots, n-1$ e $\zeta_n^k \neq \zeta_n^j$ para todo $1 \leq k < j \leq n-1$. Podemos escrever então:

$$X^n - 1 = \prod_{j=0}^{n-1}(X - \zeta_n^j).$$

Observe que $\varepsilon = \zeta_n^k$ é uma raiz primitiva da unidade se, e só se, $(k,n) = 1$. Além disso o número de raízes n-ésimas primitivas da unidade é dado por $\varphi(n)$ onde φ é a função de Euler. Portanto o seguinte polinômio é o polinômio de menor grau que possui todas as raízes n-ésimas primitivas da unidade:

$$\Phi_n(X) = \prod_{1\leq k\leq n-1\ ,\ (k,n)=1}(X - \zeta_n^k).$$

Mais ainda, temos que $X^n - 1 = \prod_{m|n}\Phi_m(X)$.

É um exercício mostrar que se p é primo e $r \geq 1$ então:

$$\begin{aligned}
\Phi_p(X) &= \tfrac{X^p-1}{X-1} = X^{p-1} + X^{p-2} + \cdots + X + 1 \\
\Phi_{p^r}(X) &= \tfrac{X^{p^r}-1}{X^{p^{r-1}}-1} = X^{p^{r-1}(p-1)} + X^{p^{r-1}(p-2)} + \cdots + X^{p^{r-1}} + 1.
\end{aligned}$$

Observe que para todo $k \geq 1$, $\Phi_k(X)$ é mônico com coeficientes inteiros e que se $k > 1$ temos que $\Phi_k(0) = 1$. Também pode-se provar que $\Phi_k(1)$ é igual à p se k é uma potência de p e é igual à 1 caso

contrário. Finalmente pode-se mostrar que $|\Phi_k(a)| > 1$ para todo $a > 1$.

Iremos usar as seguintes identidades. Se p é primo e divide n então $\Phi_{pm}(X) = \Phi_m(X^p)$ e se p não divide m e $r \geq 1$ então:

$$\Phi_{mp^r}(X) = \frac{\Phi_m(X^{p^r})}{\Phi_m(X^{p^{r-1}})}.$$

Além disso, utilizaremos um resultado devido a Legendre que diz que os seguintes conjuntos formados por primos são iguais:

$$\begin{aligned} E_1 &= \{p\,; p|a^n - 1 \text{ ,mas } p \nmid a^m - 1,\ \forall 1 \leq m \leq n-1\} \\ E_2 &= \{p\,; p|\Phi_n(a) \text{ e } p \equiv 1 \pmod{n}\} \\ E_3 &= \{p\,; p \nmid n \text{ e } p|\Phi_n(a)\}. \end{aligned}$$

Com este resultado, podemos provar o teorema de Dirichlet no caso em que $a = 1$. Sejam $p_i \equiv 1 \pmod{b}$ primos com $i = 1, \ldots, r$ e defina $N = bp_1 \ldots p_r$. Temos então que $|\Phi_b(N)| > 1$.

Tome p um primo que divide $\Phi_b(N)$. Pelas identidades acima citadas, vemos que $\Phi_b(N) \equiv \Phi_b(0) \equiv 1 \pmod{N}$. Portanto p não divide N, logo p não divide b. Pelo resultado de Legendre, segue que $p \equiv 1 \pmod{b}$ e $p \neq p_i$ com $i = 1, \ldots, r$. Repetindo o processo, encontramos infinitos primos na progressão aritmética $\{1 + kb\}$ com $k \geq 1$.

1.9.2 Prova da proposição 1.4.2

Primeiramente, iremos estudar a analiticidade da função zeta.

Proposição 1.9.1. *1 é o único polo da função zeta de Riemann em $Re(s) > 0$, ele é simples e tem resíduo 1. De fato, em $Re(s) > 0$ temos a expansão:*

$$\zeta(s) = \frac{1}{s-1} + 1 + s\int_1^\infty \frac{([x] - x)}{x^{1+s}}dx.$$

Demonstração. Seja $P(x) = [x]$. Então, em $Re(s) > 2$, temos que $\sum_{n\geq 1} \frac{P(n)}{n^s}$ e $\sum_{n\geq 1} \frac{P(n-1)}{n^s}$ convergem e, obviamente, $\int_1^\infty P(x)x^{-1-s}dx$ é analítica em $Re(s) > 1$. Invocamos então o seguinte lema:

Lema 1.9.1. *Sejam* $f(s) = \sum_{n\geq 0} \frac{a_n}{n^s}$, *em* $Re(s) > a$, *uma função meromorfa e* $P(x) = \sum_{n\leq x} a_n$ *tal que* $\sum \frac{P(n)}{n^s}$ *e* $\sum \frac{P(n-1)}{n^s}$ *convergem em* $Re(s) > b$ *e* $\int_1^\infty P(x)x^{-1-s}dx$ *é analítica em* $Re(s) > c$. *Então:*

$$f(s) = s\int_1^\infty P(x)x^{-1-s}dx.$$

Assim temos que $\zeta(s) = s\int_1^\infty [x]x^{-1-s}dx$ em $Re(s) > 1$ (onde [x] é a função maior inteiro menor que x). Por outro lado:

$$s\int_1^\infty x.x^{-1-s}dx = s\int_1^\infty \frac{1}{x^s}dx = \frac{s}{1-s}(x^{-s+1})|_1^\infty = 1 + \frac{1}{s-1}.$$

Em particular $\zeta(s) = \frac{1}{s-1} + 1 + s\int_1^\infty ([x] - x)x^{-1-s}dx$, e o resultado segue pois a integral converge em $Re(s) > 0$.

A prova do lema segue das seguintes igualdades em $Re(s) > \max\{a, b\}$:

$$\begin{aligned} f(s) &= \sum \frac{a_n}{n^s} = \sum \frac{P(n) - P(n-1)}{n^s} = \sum P(n)(\frac{1}{n^s} - \frac{1}{(n+1)^s}) \\ &= \sum sP(n)\int_n^{n+1} x^{-1-s}dx = s\int_1^\infty P(x)x^{-1-s}. \end{aligned}$$

Agora é usar continuação analítica. □

A seguir iremos obter uma região livre de zeros para a função zeta. Esta região é relativamente simples de obter com as representações anteriores:

Proposição 1.9.2. $\zeta(s) \neq 0$ *em* $Re(s) \geq 1$.

Demonstração. Vamos considerar primeiro o caso em que $Re(s) \geq \sigma > 1$. Então pela fórmula de Euler temos:

$$\frac{1}{|\zeta(s)|} = \prod_p |1 - p^{-s}| \leq \prod_p (1 + \frac{1}{|p^s|}) \leq \prod_p (1 + \frac{1}{p^\sigma}).$$

Agora este último produtório é convergente pois:

$$\begin{aligned}
\log\prod_{p_1}^{p_n}(1+\frac{1}{p^\sigma}) &= \sum_{p_1}^{p_n}\log(1+\frac{1}{p^\sigma}) = \sum_{p_1}^{p_n}\sum_{m\geq 1}\frac{(-1)^{m+1}p^{-m\sigma}}{m} \\
&\leq \sum_{p_1}^{p_n}p^{-\sigma}+\sum_{p_1}^{p_n}\sum_{m\geq 2}p^{-\sigma m} \\
&\leq \sum_{p_1}^{p_n}p^{-\sigma}+\sum_{p_1}^{p_n}\frac{p^{-2\sigma}}{1-p^{-\sigma}} \\
&\leq 2\sum_{p_1}^{p_n}p^{-\sigma} < \infty.
\end{aligned}$$

De fato, o mesmo argumento prova que se $a_i \geq 0$ e $\sum a_i < \infty$ então $\prod(1+a_i) < \infty$. Portanto $\zeta(s) \neq 0$ em $Re(s) > 1$.

Na reta $Re(s) = 1$ iremos usar a seguinte identidade trigonométrica: $3 + 4\cos\theta + \cos 2\theta = 2(1+\cos\theta)^2 \geq 0$. Vamos supor por absurdo que existe um b tal que $\zeta(1+ib) = 0$. Considere a função $\phi(s) = \zeta^3(s)\zeta^4(s+ib)\zeta(s+2ib)$. Note que $s = 1$ é um zero de ϕ (a ordem do pólo cancela com a ordem do zero), portanto $\lim_{s\to 1}\log|\phi(s)| = -\infty$.

Por outro lado, se $s > 1$ é real, temos que, para alguma sequência a_n, vale

$$\begin{aligned}
\log|\zeta(s+it)| &= Re(\log\zeta(s+it)) = Re\log(\prod_p(1-p^{-s-it})^{-1}) = \\
&= -Re(\sum\log(1-p^{-s-it})) = Re(\sum\frac{p^{(-s-it)m}}{m}) \\
&=: Re(\sum a_n n^{-s-it}).
\end{aligned}$$

Logo:

$$\begin{aligned}
\log|\phi| &= 3Re(\sum a_n n^{-s}) + 4Re(\sum a_n n^{-s-ib}) + Re(\sum a_n n^{-s-2ib}) \\
&= Re(\sum a_n n^{-s}(3+4n^{-ib}+n^{-2ib})) \\
&= Re(\sum a_n n^{-s}(3+4e^{-ib\log n}+e^{-2ib\log n})) \\
&= \sum a_n n^{-s}(3+4\cos(b\log n)+\cos(2b\log n)) \geq 0.
\end{aligned}$$

Esta contradição finaliza a prova. □

Vemos também que o argumento acima implica que a função $-\frac{\zeta'(s)}{\zeta(s)}$ tem um pólo em $s = 1$ do tipo $\frac{1}{s-1}$ (o qual é único em $Re(s) \geq 1$).

Outra maneira de achar uma região livre de zeros é a seguinte: usando a expansão do logaritmo em (1.4.1), temos que (pela mudança de variável $n = dm$) em $Re(s) > 1$ vale:

$$\begin{aligned}\sum_n \frac{\log n}{n^s} &= \sum_n \sum_{d|n} \frac{\Lambda(d)}{d^s} = \sum_n \sum_{d|n} \frac{\Lambda(d)}{(dm)^s} \\ &= \sum \frac{\Lambda(d)}{d^s} \sum_m \frac{1}{m^s} = \sum \frac{\Lambda(d)}{d^s}.\zeta(s)\end{aligned}$$

Por outro lado:

$$\begin{aligned}\frac{d}{ds}\zeta(s) &= \sum \frac{d}{ds} m^{-s} = \sum \frac{d}{ds}(e^{-s\log m}) = \\ &= \sum m^{-s}(-\log m) = -\sum \frac{\log m}{m^s},\end{aligned}$$

donde temos a equação:

$$\sum \frac{\Lambda(d)}{d^s} = -\frac{\zeta'(s)}{\zeta(s)}.$$

Agora somando por partes temos que:

$$\begin{aligned}\sum_1^N \frac{1}{n^s} &\approx \frac{n}{n^s}|_1^N + \sum_1^N \frac{s}{n^s}, \\ (1-s)\sum_1^N \frac{1}{n^s} &\approx \frac{1}{N^{s-1}} - 1, \\ \sum \frac{1}{n^s} &= \frac{1}{s-1} + O(1).\end{aligned}$$

Novamente somando por partes:

$$\begin{aligned}\sum_1^N \frac{\log n}{n^s} &\approx \frac{\log n}{n^{s-1}}|_1^N - \sum_1^N \frac{n^s - sn^s \log n}{n^2 s},\\ (1-s)\sum_1^N \frac{\log n}{n^s} &\approx \frac{\log N}{N^{s-1}} - \frac{1}{s-1} + O(1),\\ \sum \frac{\log n}{n^s} &= \frac{1}{(s-1)^2} + O(1).\end{aligned}$$

Em particular não existem zeros em $Re(s) > 1$ com $s \approx 1$ e também obtemos:

$$\sum \frac{\Lambda(d)}{d^s} = \frac{1}{s-1} + O(1).$$

Veremos agora algumas estimativas sobre ζ que serão úteis no Capítulo 3. Já sabemos que $\zeta(s) = \frac{1}{s-1} + \sum_{n=1}^{\infty}(\frac{1}{n^s} - \int_0^1 \frac{dx}{(n+x)^s})$ em $\{\text{Re } s > 0\}$, e $\text{Re}(\log(\zeta(\sigma+2it)\zeta(\sigma+it)^4\zeta(\sigma)^3)) \geq 0$, donde $|\zeta(\sigma+2it)\zeta(\sigma+it)^4\zeta(\sigma)^3| \geq 1$, para $\sigma \geq 1$, $t \in \mathbb{R}$. Derivando a expressão para $\zeta(s)$, obtemos

$$\zeta'(s) = -\frac{1}{(s-1)^2} + \sum_{n=1}^{\infty}(\int_0^1 \frac{\log(n+x)dx}{(n+x)^s} - \frac{\log n}{n^s}) \text{ em } \{\text{Re } s > 0\}.$$

Suponha que $10 \geq \text{ Re } s \geq 1 - \frac{b}{\log(|t|+2)}$, onde $t = \text{ Im } s$ e b é uma constante com $0 < b < 1/2$. Temos então

$$\begin{aligned}|\zeta(s) - \frac{1}{s-1}| &\leq \sum_{n=1}^{\infty} |\frac{1}{n^s} - \int_0^1 \frac{dx}{(n+x)^s}|\\ &= \sum_{1\leq n\leq |t|} |\frac{1}{n^s} - \int_0^1 \frac{dx}{(n+x)^s}| + \sum_{n>|t|} |\frac{1}{n^s} - \int_0^1 \frac{dx}{(n+x)^s}|\\ &\leq 2\sum_{1\leq n\leq |t|} \frac{1}{n^{\text{Re}s}} + \sum_{n>|t|} \frac{|s|}{n^{\text{Re } s+1}},\end{aligned}$$

pois $\frac{1}{n^s} - \frac{1}{(n+x)^s} = f(0) - f(x) = -f'(d)\cdot x$, para algum $d \in (0,x)$, onde $f(y) = (n+y)^{-s}$ satisfaz $f'(y) = -s(n+y)^{-s-1} \Rightarrow |f'(y)| \leq \frac{|s|}{n^{\mathrm{Re}s+1}}$, $\forall\, y \in (0,1)$. Assim,

$$\begin{aligned}\Big|\zeta(s) - \frac{1}{s-1}\Big| &\leq 2\sum_{1\leq n\leq |t|} \frac{1}{n^{1-b/\log(|t|+2)}} + \frac{|s|}{|t|^{1-b/\log(|t|+2)}} \\ &\leq 2|t|^{b/\log(|t|+2)} \sum_{1\leq n\leq |t|} \frac{1}{n} + \frac{|s|\cdot |t|^{b/\log(|t|+2)}}{|t|} \\ &= \mathcal{O}(\log(|t|+2)).\end{aligned}$$

Temos também

$$\begin{aligned}\Big|\zeta'(s) + \frac{1}{(s-1)^2}\Big| &\leq \sum_{1\leq n\leq |t|} \Big|\int_0^1 \frac{\log(n+x)dx}{(n+x)^s} - \frac{\log n}{n^s}\Big| \\ &\quad + \sum_{n>|t|} \Big|\int_0^1 \frac{\log(n+x)dx}{(n+x)^s} - \frac{1}{n^s}\Big| \\ &\leq 2\sum_{1\leq n\leq |t|} \frac{\log n}{n^{\mathrm{Re}\ s}} + \sum_{n>|t|} \frac{|s|\log n}{n^{\mathrm{Re}\ s+1}} \\ &= \mathcal{O}\big(|\log t|^2 \quad + \frac{|s|\log|t|}{|t|^{\mathrm{Re}\ s}}\big) \\ &= \mathcal{O}((\log(|t|+2))^2).\end{aligned}$$

Seja agora $Z = \big\{z \in \mathbb{C} \mid 10 \geq \ \mathrm{Re}\ z \geq 1 - \frac{\beta}{(\log(|\mathrm{Im}\ z|+2))^9}\big\}$, onde β é uma constante pequena. Temos $|\zeta(\sigma+2it)\zeta(\sigma+it)^4\zeta(\sigma)^3| \geq 1$ para $\sigma \geq 1$, donde, como $|\zeta(\sigma+2it)| = \mathcal{O}(\log(|t|+2))$, para $\sigma + it \in Z$, escolhendo $\sigma = 1 + c/(\log(|t|+2))^9$, onde $c > 0$ é uma constante pequena, $|\zeta(\sigma)| = \mathcal{O}(c^{-1}(\log(|t|+2))^9)$, donde temos $|\zeta(\sigma+it)^{-4}| \leq |\zeta(\sigma)|^3|\zeta(\sigma+2it)| = \mathcal{O}(c^{-3}(\log(|t|+2))^{28})$, e portanto $|\zeta(\sigma+it)^{-1}| = \mathcal{O}(c^{-3/4}(\log(|t|+2))^7)$. Como $|\zeta'(x+it)| = \mathcal{O}((\log(|t|+2))^2)$ para $x \in [1,\sigma]$, segue que $|\zeta(1+it)^{-1}| = \mathcal{O}((\log(|t|+2))^7)$, se tomarmos a constante $c > 0$ suficientemente pequena. De fato, $\zeta(\sigma+it) \geq Ac^{3/4}(\log(|t|+2))^{-7}$ para uma certa

constante positiva A independente de c. Assim, para

$$1 - \frac{c}{(\log(|t|+2))^9} \leq a \leq 1 + \frac{c}{(\log(|t|+2))^9}$$

temos

$$\begin{aligned} |\zeta(a+it) - \zeta(\sigma+it)| &\leq |a-\sigma| \cdot \max\{|\zeta'(x+it)|, a \leq x \leq \sigma\} \\ &= \mathcal{O}(\frac{c}{(\log(|t|+2))^9} \cdot (\log(|t|+2))^2) \\ &= \mathcal{O}(c(\log(|t|+2))^{-7}), \end{aligned}$$

donde

$$\begin{aligned} |\zeta(a+it)| &\geq Ac^{3/4}(\log(|t|+2))^{-7} - \mathcal{O}(c(\log(|t|+2))^{-7}) \\ &> \frac{1}{2}Ac^{3/4}(\log(|t|+2))^{-7}, \end{aligned}$$

para c suficientemente pequeno. Se, por outro lado, $\sigma < d \leq 10$, de $|\zeta(d+2it)\zeta(d+it)^4\zeta(d)^3| \geq 1$, segue que $|\zeta(d+it)^{-4}| \leq |\zeta(d)|^3|\zeta(d+2it)| \leq |\zeta(\sigma)|^3|\zeta(d+2it)| = \mathcal{O}((\log(|t|+2))^{28})$, pois $|\zeta(d+2it)| = \mathcal{O}(\log(|t|+2))$. Assim, $|\zeta(d+it)^{-1}| = \mathcal{O}((\log(|t|+2))^7)$, e segue que a estimativa para $|\zeta^{-1}|$ vale em toda a região Z.

Finalmente, se $0 < \lambda < 1$ e $\sigma \geq \lambda$, temos, como antes, para $s = \sigma + it$, $\left|\zeta(s) - \frac{1}{s-1}\right| \leq \sum_{n=1}^{\infty} \left|\frac{1}{n^s} - \int_0^1 \frac{dx}{(n+x)^s}\right| \leq 2\sum_{1\leq n\leq |t|} \frac{1}{n^\sigma} + \sum_{n>|t|} \frac{|s|}{n^{\sigma+1}} \leq 2\sum_{1\leq n\leq |t|} \frac{1}{n^\lambda} + \mathcal{O}(|s||t|^\sigma) = \mathcal{O}(|t|^{1-\lambda}) + \mathcal{O}(|t|^{1-\sigma}) = \mathcal{O}(|t|^{1-\lambda})$. Em particular, se Re $s \geq 3/4$, $\left|\zeta(s) - \frac{1}{s-1}\right| = \mathcal{O}(|t|^{1/4})$.

1.9.3 Prova do teorema 1.4.2

Nesta seção, iremos fixar uma $f \in L^1_{loc}([1,\infty))$, $f \geq 0$, não decrescente com $f(x) = O(x)$ e denotaremos por g sua transformada de Mellin. Primeiramente vamos provar que g *é analítica em* $Re(s) > 1$. De fato, fixe s tal que $Re(s) > \sigma > 1$ e tome um A grande tal que se $x > A$ então $|f(x)| \leq C|x|$. Temos que:

$$\left|\int_\lambda^\infty f(x)x^{-1-s}\right| \leq \int_\lambda^\infty Cx^{-\sigma} \leq \frac{C}{\sigma-1}\lambda^{1-\sigma}.$$

Isto diz que $\lim_{\lambda\to\infty} \int_1^\lambda f(x)x^{-1-s} = \int_1^\infty f(x)x^{-1-s}$. Logo g é analítica em $Re(s) > 1$.

Lembramos que *se* $F \in L^1_{loc}(0,\infty)$ *é limitada, então a transformada de Laplace* $L(z) = \int_0^\infty F(t)e^{-zt}dt$ *é analítica em* $Re(z) > 0$. Agora, um ponto importante é que *se ela se estende para* $Re(z) = 0$ *então* $L(0) = \int_0^\infty F(t)dt$. Provaremos isto mais tarde.

Vamos usar este fato para a função $F(t) = e^{-t}f(e^t) - c$, com $t > 0$. Ora, as hipóteses sobre f garantem que ela é limitada e está em $L^1_{loc}(0,\infty)$. Por outro lado, a mudança de variável $x = e^t$ diz que a transformada de Laplace de f é:

$$\begin{aligned} L(z) &= \int_0^\infty (e^{-t}f(e^t) - c)e^{-zt}dt = \int_1^\infty x^{-2-z}f(x)dx - (\frac{cx^{-z}}{-z})|_1^\infty \\ &= \frac{g(z+1)}{z+1} - \frac{c}{z} = \frac{1}{z+1}(g(z+1) - \frac{c}{z} - c). \end{aligned}$$

Logo, pela hipótese de extensão de f, temos que L se estende à $Re(z) = 0$ e portanto $L(0) = \int_1^\infty \frac{f(x)-cx}{x^2}dx$. Como consequência, temos que $c > 0$ (do contrário a integral seria infinita pois $f \geq 0$).

Vamos supor, por absurdo, que existe um $\delta > 0$ tal que $\limsup \frac{f(x)}{x} - c > 2\delta > 0$. Tome $\rho = \frac{c+2\delta}{c+\delta} > 1$ e $y_n \to \infty$ uma sequência tal que $f(y_n) > (c + 2\delta)y_n$. Como f é não decrescente, então para todo $y_n < x < \rho y_n$ vale:

$$f(x) \geq f(y_n) > (c + 2\delta)y_n > (c + \delta)x.$$

Portanto, temos que $\psi(x) := \frac{f(x)-cx}{x^2} > \frac{\delta}{x}$. Isto implica que:

$$\int_{y_n}^{\rho y_n} \psi(x)dx \geq \int_{y_n}^{\rho y_n} \frac{\delta}{x}dx = \delta \log \rho > 0.$$

Fixando $\varepsilon < \frac{\delta}{2}\log\rho$, segue que se a é grande temos (por convergência da integral) que: $|\int_a^\infty \psi(x)| < \varepsilon$. Podemos tomar $a \geq y_{n_0}$ para algum n_0, daí temos que:

$$\delta \log \rho < |\int_{y_{n_0}}^{\rho y_{n_0}} \psi(x)dx| \leq |\int_{y_{n_0}}^\infty \psi(x) - \int_{\rho y_{n_0}}^\infty \psi(x)| < 2\varepsilon < \delta \log \rho.$$

Este absurdo implica $\limsup \frac{f(x)}{x} \leq c$. Um argumento análogo dá a desigualdade desejada para o $\liminf$.

Faltou então provar a afirmação sobre a transformada de Laplace, isto é que a integral $\int_0^\infty F(t)$ converge e é igual à $L(0)$ se a transformada de Laplace se estende analiticamente para $Re(z) = 0$. Usando um reescalonamento, podemos supor que $|F| \leq 1$. Vamos considerar as integrais truncadas $L_\lambda(z) = \int_0^\lambda F(t)e^{tz}$, as quais definem funções analíticas em todo o plano complexo, e provar que $\lim_{\lambda\to\infty} L_\lambda(0) = L(0)$.

Seja então $\varepsilon > 0$ pequeno e tome $R = \varepsilon^{-1}$. Por hipótese, é claro que L tem continuação analítica em uma vizinhança de $\{Re(z) \geq 0\}$, logo existe um $\delta > 0$ tal que L é analítica em $B = D(0, R) \cap \{Re(z) \geq -\delta\}$. Se $W = \partial B$, a fórmula de Cauchy diz que:

$$L(0) - L_\lambda(0) = \frac{1}{2\pi i}\int_W \frac{L(z) - L_\lambda(z)}{z}dz.$$

Agora usamos o seguinte truque: se ψ é analítica então a fórmula de Cauchy diz que $2\pi i\psi(0) = \int_W \frac{\psi(z)e^{\lambda z}}{z}dz$ e $0 = \int_W \frac{\psi(z)e^{\lambda z}z^2}{R^2 z}dz$. Somando as duas igualdades e aplicando-as a função $L - L_\lambda$ temos que:

$$L(0) - L_\lambda(0) = \int_W (L(z) - L_\lambda(z))e^{\lambda z}(\frac{1}{z} + \frac{z}{R2})dz. \qquad (1.9.1)$$

Vamos denotar por $I_\phi(z) = \phi(z)e^{\lambda z}(\frac{1}{z} + \frac{z}{R^2})$.

Agora, vamos separar essa integral em regiões. Seja $W^+ := W \cap \{Re(z) > 0\}$ e $W^- := W \cap \{Re(z) < 0\}$. Como I_L é analítica em W, ela é limitada por uma constante C em W. Além disso, existe um $\gamma < \delta$ tal que em $W_2^- := W \cap \{-\gamma \leq Re(z) < 0\}$ temos $\int_{W_2^-} |dz| < \frac{2\pi\varepsilon}{C}$. Considere então $W_1^- := W \cap \{Re(z) < \gamma\}$ e $W_*^- := \{Re(z) < 0\} \cap \{|z| = R\}$. Observe que pela analiticidade de L_λ temos que $\int_{W^-} I_{L_\lambda} = \int_{W_*^-} I_{L_\lambda}$.

Podemos decompor a integral (1.9.1) como:

$$2\pi i(L(0) - L_\lambda(0)) = \int_{W^+} I_{L-L_\lambda}(z) + \int_{W_1^-} I_L(z) + \int_{W_2^-} I_L(z) - \int_{W_*^-} I_{L_\lambda}.$$

Passando o módulo nessa igualdade e usando a notação $x = Re(z)$ e $|z| = R$ temos que:

- A primeira integral é dominada por $\int_{W^+} e^{\lambda x}\frac{e^{-\lambda x}}{x}\frac{2x}{R^2}|dz| = \frac{1}{R} = 2\pi\varepsilon$.

- A segunda integral é dominada por $2\pi B e^{-\lambda\delta_1} \int_{W_1^-} |dz|$.
- A terceira integral é dominada por $\int_{W_2^-} e^{\lambda x} B|dz| < 2\pi\varepsilon$.
- A quarta integral é dominada por $\int_{W_*^-} e^{\lambda x} \frac{e^{\lambda x}}{|x|} \frac{2|x|}{R^2} |dz| = \frac{1}{R} = 2\pi\varepsilon$.

Logo $|L(0) - L_\lambda(0)| \leq 3\varepsilon + \frac{B}{2\pi} \int_{W_1^-} |dz| e^{-\lambda\delta_1}$, como queríamos demonstrar.

1.9.4 Prova do teorema 1.5.3

A primeira observação é que se para algum k e ε a primeira parte do teorema vale para x então ela vale para uma vizinhança inteira de x e portanto para algum ponto de Z.

Em seguida, usando o lema de Zorn, podemos supor que X é minimal, isto é X não possui nenhum subconjunto Y próprio fechado tal que $T(Y) \subset Y$. Em particular, os conjuntos $\{T^m(x)\}_{m=0}^{\infty}$ são densos em X, o que mostra a afirmação para $k = 1$ (pois, por densidade, existe um $n \in N$ tal que $d(T^n(x), x) < \varepsilon$).

A prova seguirá por indução. Suponha que o teorema vale para algum $k \geq 1$, isto é, para todo $\varepsilon > 0$ existe $x \in X$ e $n \in N$ tal que $d(T^{in}(x), x) < \varepsilon$ para $i = 1, \ldots, k$. Afirmamos que o conjunto de tais pontos é denso em X.

De fato, seja $U \subset X$ um aberto e $B \subset U$ uma bola de raio menor que ε. Vamos definir $B_m = (T^m)^{-1}(B)$ de modo que estes conjuntos formam uma cobertura de X (pela minimalidade de X). Por compacidade temos uma subcobertura finita $\{B_{m_1}, \ldots, B_{m_r}\}$. Seja $\delta > 0$ um número de Lebesgue desta cobertura, ou seja, um número tal que toda bola de raio δ está contida em algum aberto desta cobertura. Tome x e n tais que $d(T^{in}(x), x) < \delta$ para $i = 1, \ldots, k$ e D a bola de centro x e raio δ. Então existe j tal que $D \subset B_{m_j}$, em particular $T^{m_j}(D) \subset B$. Ou seja que $T^{m_j}(T^{in}(x))$ pertencem a bola de raio ε centrada em $T^{m_j}(x) \in U$. Isto prova a densidade.

Vamos voltar agora a prova do teorema. Fixe $\varepsilon > 0$. Pela hipótese de indução existem x e n_0 tais que $d(T^{in_0}x_0, x_0) < \varepsilon/2$ para $i = 1, \ldots, k$. Tomando x_1 tal que $T^{n_0}(x_1) = x_0$, temos $d(T^{(i+1)n_0}x_1, x_0) < \varepsilon/2$ para $i = 1, \ldots, k$. Portanto segue que $d(T^{in_0}(x_1), x_0) < \varepsilon/2$ para $i = 1, \ldots, k+1$.

Por continuidade, existe $\varepsilon_1 < \varepsilon$ tal que se $d(y,x) < \varepsilon_1$ então $d(T^{in_0}(y), x) < \varepsilon/2$ para $i = 1, \ldots, k+1$. Pela hipótese de indução novamente, existe y_1 tal que $d(y_1, x_1) < \varepsilon_1$ e n_1 tal que $d(T^{in_1}(y_1), y_1) < \varepsilon_1/2$ para $i = 1, \ldots, k$. Por desigualdade triangular temos que:

$$d(T^{in_0}(T^{(i-1)n_1}(y_1)), x_0) < \varepsilon_2 \text{ para } i = 1, \ldots, k+1.$$

Procedendo desta maneira (tomando x_2 tal que $T^{n_1}(x_2) = y_1$) encontramos pontos $x_2, x_3, \cdots \in X$ e naturais $n_2, n_3, \ldots$ tais que para todo l temos:

$$\begin{aligned} d(T^{in_{l-1}}(x_l), x_{l-1}) &< \varepsilon/2 \\ d(T^{i(n_{l-1}+n_{l-2})}(x_l), x_{l-2}) &< \varepsilon/2 \\ &\cdots \\ d(T^{i(n_{l-1}+\cdots+n_0)}(x_l), x_0) &< \varepsilon/2 \text{ para } i = 1, \ldots, k+1. \end{aligned}$$

Por compacidade, existem $l > m$ tal que $d(x_l, x_m) < \varepsilon/2$. Por desigualdade triangular temos que:

$$d(T^{i(n_{l+1}+\cdots+n_m)}(x_l), x_l) < \varepsilon \text{ , para } i = 1, \ldots, k+1.$$

Logo, basta tomar $x = x_l$ e $n = n_{l-1} + \cdots + n_m$ para finalizar a prova do teorema.

1.9.5 O exemplo de F. Behrend

Conforme anunciamos na observação 1.7.1, primeiramente vamos construir exemplos de subconjuntos S do conjunto dos inteiros não-negativos $\leq N$ sem nenhuma progressão aritmética de tamanho 3 e com cardinalidade $|S| \geq N^{1-\frac{2\sqrt{2\log 2}+\varepsilon}{\sqrt{\log N}}}$; em seguida, adaptaremos esta técnica para estudar o comportamento da função $c(3, \delta)$.

Dados $d \geq 2$, $n \geq 2$ e $k \leq n(d-1)^2$, considere $S_k(n,d)$ o conjunto de todos os números inteiros da forma

$$x = a_1 + a_2(2d-1) + \cdots + a_n(2d-1)^{n-1}$$

cujos dígitos a_i na base $(2d-1)$ estão sujeitos as restrições

$$0 \leq a_i < d \quad \text{e} \quad \|x\|^2 := a_1^2 + \cdots + a_n^2 = k.$$

Note que $S_k(n,d)$ *não* contém progressões aritméticas de tamanho 3: com efeito, caso existissem $x, x', x'' \in S_k(n,d)$ tais que $x + x' = 2x''$, então

$$\|x + x'\| = \|2x''\| = 2\sqrt{k}$$

e

$$\|x\| + \|x'\| = 2\sqrt{k}.$$

Logo, como a igualdade na desigualdade triangular $\|x + x'\| \leq \|x\| + \|x'\|$ só pode ocorrer quando os vetores $(a_1, \ldots, a_n)$ e $(a'_1, \ldots, a'_n)$ são proporcionais, vemos que $x = x' = x''$ (porque estes vetores tem normas iguais por hipótese).

Por outro lado, existem d^n vetores $(a_1, \ldots, a_n)$ satisfazendo a restrição $0 \leq a_i < d$ e $n(d-1)^2 + 1$ valores possíveis para k. Consequentemente, para algum $k = K_0$, $S_k(n,d)$ deve ter cardinalidade ao menos

$$\frac{d^n}{n(d-1)^2+1} > \frac{d^{n-2}}{n}.$$

Como todos os elementos de $S_k(n,d)$ possuem módulo $< (2d-1)^n$, se definirmos

$$\nu(N) := \max\{|S| \,;\, S \subset [1,N],\, S \text{ sem nenhuma 3-PA}\},$$

obtemos que

$$\nu((2d-1)^n) > d^{n-2}/n.$$

Agora, fixado $\varepsilon > 0$ e dado N grande, escolhemos $n = \lfloor \sqrt{\frac{2\log N}{\log 2}} \rfloor$ e d satisfazendo

$$(2d-1)^n \leq N < (2d+1)^n,$$

de maneira que

$$\begin{aligned}
\nu(N) \geq \nu((2d-1)^n) &> \frac{d^{n-2}}{n} > \frac{(N^{1/n}-1)^{n-2}}{2^{n-2}n} \\
&= \frac{N^{1-(2/n)}}{2^{n-2}n}(1 - N^{-1/n})^{n-2} \\
&> \frac{N^{1-(2/n)}}{2^{n-1}n} = N^{1-(2/n)-\frac{\log n}{\log N}-\frac{(n-1)\log 2}{\log N}} \\
&> N^{1-\frac{2\sqrt{2\log 2}+\varepsilon}{\sqrt{\log N}}}.
\end{aligned}$$

Agora pretendemos modificar ligeiramente o raciocínio anterior para estudar o comportamento da função $c(3,\delta)$: fixamos $d, n \geq 1$ inteiros (a serem escolhidos mais tarde) e definimos $\phi : \{1,\dots,N\} \to \{0,\dots,2d-1\}^n$ por

$$\phi(x) := (\lfloor x/(2d-1)^i \rfloor \mathrm{mod}(2d-1))_{i=0}^{n-1}.$$

Para cada k entre 1 e $n(d-1)^2$, considere novamente os conjuntos

$$S_k(n,d) := \{(x_1,\dots,x_n) \in \{0,\dots,d-1\}^n \,:\, x_1^2+\cdots+x_n^2 = k\}$$

e defina $A_k(n,d) := \phi^{-1}(S_k(n,d))$. Conforme já sabemos, $S_k(n,d)$ é livre de 3-PA (*exceto* as 3-PAs triviais $\{x,x,x\}$). Isto implica que $A_k(n,d)$ só pode conter progressões aritméticas $(n, n+r, n+2r)$ onde r é um múltiplo de $(2d-1)^n$. Em particular, o número máximo de 3-PAs em $A_k(n,d)$ é $N^2/(2d-1)^n$. Por outro lado, quando $\phi(x) \in \{0,\dots,d-1\}^n$, a probabilidade de x pertencer a $A_k(n,d)$ é $\frac{1}{n(d-1)^2+1}$. Logo, temos a seguinte cota inferior para a cardinalidade de $A_k(n,d)$:

$$|A_k(n,d)| \geq \frac{c}{nd^2} 2^{-n} N.$$

Tomando $n = c\log(1/\delta)$ e $d = \delta^{-c}$, obtemos que, para algum k, o conjunto $A_k(n,d)$ satisfaz $|A_k(n,d)| \geq \delta^c N$ e o número máximo de 3-PAs em $A_k(n,d)$ é $\delta^{c\log(1/\delta)} N^2$. Em outras palavras, $c(3,\delta) \leq \delta^{c\log(1/\delta)}$.

Capítulo 2

Teorema de Green-Tao-Szemerédi

2.1 Introdução

O principal objetivo deste capítulo é apresentar as idéias da prova do teorema de Green e Tao.

Grosseiramente falando, a prova deste teorema consiste em dois passos:

- generaliza-se o teorema de Szemerédi para o contexto de *medidas pseudo-aleatórias* obtendo-se assim o teorema de Green-Tao-Szemerédi (veja a seção 2.2 para mais detalhes);
- prova-se a existência de medidas pseudo-aleatórias nos primos (ao longo das linhas dos recentes resultados de Goldston-Yıldırım).

Uma vez que estes dois fatos já estejam provados, veremos que o teorema de Green e Tao segue diretamente (veja a seção 2.2).

Porém, antes de entrar (na seção 2.4) nos detalhes do esboço delineado acima , pretendemos motivar os conceitos e táticas da prova do teorema de Green-Tao-Szemerédi através do esquema de prova do teorema de Roth (o qual corresponde ao caso particular $k = 3$

no teorema de Szemerédi) na seção 2.3 (enquanto que iremos deixar as discussões relativas aos resultados de Goldston-Yıldırım para o capítulo 3).

A organização deste capítulo será assim:

- na seção 2.2 apresentaremos mais detalhadamente o esboço da prova do teorema de Green-Tao; em particular, iremos enunciar precisamente os teoremas de Green-Tao-Szemerédi e Goldston-Yıldırım; finalmente, demonstraremos o teorema de Green-Tao *assumindo* estes dois resultados.

- na seção 2.3 esboçaremos a prova do teorema de Roth sobre a existência de uma quantidade infinita de progressões aritméticas de tamanho 3 (i.e., 3-PA) em conjuntos de densidade positiva, o qual servirá de motivação para a prova do teorema de Green-Tao-Szemerédi.

- finalizando este capítulo, na seção 2.4, faremos a demonstração do teorema de Green-Tao-Szemerédi.

Fechando esta introdução, observamos que ao fim deste capítulo o teorema de Green e Tao estará demonstrado exceto pelos resultados de Goldston e Yıldırım, os quais deixaremos para discutir apenas no próximo capítulo.

2.2 Estratégia da prova do teorema de Green e Tao

Durante o resto deste capítulo, nós iremos fixar $k \geq 3$ o tamanho da progressão aritmética (PA) de primos que desejamos encontrar e $N := |\mathbb{Z}_N|$ será um número primo (grande) de modo que os elementos $1, \ldots, N-1$ podem ser invertidos em $\mathbb{Z}_N$. Escreveremos $o(1)$ para denotar quantidades que tendem a zero quando $N \to \infty$ e $O(1)$ para denotar quantidades que ficam limitadas quando $N \to \infty$. Em certos lugares do texto, as quantidades $o(1)$ (resp., $O(1)$) tendem a zero (resp., permanecem limitadas) *dependendo de certos parâmetros* (p. ex., j, ε). Quando isto ocorrer, colocaremos os parâmetros subscritos

nas quantidades (p. ex., $o_{j,\varepsilon}(1)$). Além disso, abreviaremos quantidades da forma $O(1)X$ (resp., $o(1)X$) como $O(X)$ (resp., $o(X)$).

Com estas notações em mãos, podemos iniciar a contextualização do teorema de Green e Tao. Lembre que este teorema diz que para todo $k \geq 3$, existem infinitas k-PA (i.e., progressoẽs aritméticas de tamanho k) formadas (apenas) por números primos. Além disso, sabemos que o teorema de Szemerédi garante a existência de muitas k-PA em subconjuntos de inteiros com *densidade positiva*. Mais ainda, já sabemos também que o teorema de Szemerédi não implica o teorema de Green e Tao porque os primos possuem densidade zero. Entretanto, a idéia de Green e Tao é :

- apesar do teorema de Szemerédi não se aplicar *diretamente*, podemos *modificá-lo* para que ele funcione em subconjuntos com comportamento (aditivo) fracamente aleatório[1] (ou mais precisamente *pseudo-aleatório*); este resultado será chamado neste livro de *teorema de Green-Tao-Szemerédi*;

- isto reduz o teorema de Green-Tao a provar que o conjunto dos números primos é pseudo-aleatório, um fato que segue mais ou menos diretamente dos trabalhos de Goldston e Yıldırım.

Daqui em diante, iremos detalhar os itens discutidos acima. Para isso, começaremos com a definição de *pseudo-aleatoriedade*:

Definição 2.2.1.

- *Dizemos que* $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ *é uma* medida *se* $\mathbb{E}(\nu) = 1 + o(1)$.

- *Uma medida* $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ *satisfaz a* (m_0, t_0, L_0)-condição de formas lineares *se, para toda família de* $m \leq m_0$ *formas lineares* $\psi_i : \mathbb{Z}_N^t \to \mathbb{Z}_N$, $t \leq t_0$, *digamos* $\psi_i(x) = \sum L_{ij}x_j + b_i$, *onde* L_{ij} *são racionais com numeradores e denominadores menores que* L_0, *as* t*-uplas* $(L_{ij})_{1\leq j\leq t}$ *não são múltiplas racionais entre si e* $b_i \in \mathbb{Z}$ *quaisquer, então:*

$$\mathbb{E}(\nu(\psi_i(x)) \dots \nu(\psi_m(x))|x \in \mathbb{Z}_N^t) = 1 + o_{m_0,t_0,L_0}(1).$$

[1] O ponto de pedir aleatoriedade fraca é que desejamos depois usar o resultado para os primos.

- *Uma medida* $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ *satisfaz a* m_0-condição de correlação *se para todo* $m \leq m_0$ *existem pesos* $\tau_m : \mathbb{Z}_N \to \mathbb{R}^+$ *tais que* $\mathbb{E}(\tau_m^q) = O_{m,q}(1)$ *(condição de momentos) para todo* $1 \leq q < \infty$ *e*

$$\mathbb{E}(\nu(x+h_1)\dots\nu(x+h_m)|x \in \mathbb{Z}) \leq \sum_{i<j\leq m} \tau_m(h_i - h_j).$$

  para quaisquer $h_i \in \mathbb{Z}_N$ *(não necessariamente distintos).*

- *Uma medida* $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ *é* k-pseudo-aleatória *se* $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ *satisfaz a* $(k \cdot 2^{k-1}, 3k-4, k)$ *condição de formas lineares e a* 2^{k-1} *condição de correlação.*

A definição pode parecer estranha no início, porém ela é baseada em trabalhos de Goldston-Yıldırım, onde estuda-se majorantes para funções modificadas de von Mangoldt (que por sua vez estão intimamente ligadas aos números primos). Intuitivamente, as condições acima falam que o conjunto de inteiros no suporte de ν tem propriedades aritméticas (aditivas) fracamente aleatórias. A principal vantagem destas condições é que elas são suficientemente fracas para incluir o caso dos primos (apesar deste fato estar longe de ser trivial) e permitir o uso de uma versão do teorema de Szemerédi para o suporte destas medidas (veja o teorema 2.2.1 logo abaixo).

Munidos do conceito de pseudo-aleatoriedade, estamos aptos para enunciar um dos resultados principais do trabalho de Green e Tao [5, Theorem 3.5]:

Teorema 2.2.1 (Green-Tao-Szemerédi)**.** *Se* $k \geq 3$, $0 < \delta \leq 1$ *e* ν *é* k*-pseudo-aleatória então para toda* $f : \mathbb{Z}_N \to \mathbb{R}^+$ *tal que* $0 \leq f(n) \leq \nu(n)$ *e* $\mathbb{E}(f) \geq \delta$ *temos que:*

$$\mathbb{E}(f(n)f(n+r)\dots f(n+(k-1)r)|n, r \in \mathbb{Z}_N) \geq c(k,\delta) - o_{k,\delta}(1).$$

Observação 2.2.1. *Note que fazendo* $\nu \equiv \nu_{const} \equiv 1$ *no teorema acima obtemos automaticamente o teorema de Szemerédi como corolário.*

A prova do teorema 2.2.1 é baseada nas idéias de *Furstenberg* (de recorrência múltipla em teoria ergódica) e nas *normas de Gowers*. Por enquanto, adiaremos a demonstração do teorema 2.2.1 para a seção 2.4.

Visando aplicar o teorema 2.2.1 para finalizar a prova do teorema de Green e Tao, agora vamos ver como construir medidas (pseudo) aleatórias relacionadas aos primos. Iniciamos por relembrar a definição:

Definição 2.2.2. *A* função de von Mangoldt *é*

$$\Lambda(n) := \begin{cases} \log p \text{ } se \text{ } n = p^m \\ 0 \text{ } caso \text{ } contrário \text{ } . \end{cases}$$

Lembre que esta função está (essencialmente) suportada nos primos (pois a contribuição de potências de primos é pequena), de modo que ela funciona como uma "função característica" dos primos. Em termos dela, sabemos que o *teorema de números primos* pode ser reformulado como $\mathbb{E}(\Lambda(n)) = 1 + o(1)$. Para usar o teorema de Green-Tao-Szemerédi no contexto dos números primos, precisamos achar uma medida k-pseudoaleatória tal que $\nu(n) \geq c(k)\Lambda(n)$. Porém é sabido que tais medidas *não existem*![2]

Para contornar esse problema usa-se o "W-trick". Seja $w = w(N) \to \infty$ um parâmetro que pode crescer com N porém lentamente (i.e., $\frac{1}{w(N)} = o(1)$) e seja $W = \Pi_{p \leq w(N);\ \text{p é primo}} p$. A função de von Mangoldt modificada é:

$$\widetilde{\Lambda}(n) = \begin{cases} \frac{\phi(W)}{W} \log(Wn+1) \text{ se } Wn+1 \text{ é primo} \\ 0 \text{ caso contrário } . \end{cases}$$

Agora temos uma função que ainda vê os primos porém deixamos de ver potências e certas não-uniformidades vindas de produtos de fatores pequenos. Considere $w(n)$ com crescimento lento[3], digamos $w(n) << \log\log\log n$, de maneira que o teorema de Dirichlet diz que:

$$\frac{1}{N} \sum_{n \leq N} \widetilde{\Lambda}(n) = 1 + o(1).$$

[2]Basicamente porque os primos e a função de van Mangoldt estão concentrados, para todo $q > 1$ inteiro, nas $\phi(q)$ classes residuais $a (\text{mod } q)$ tais que $(a, q) = 1$ (onde $\phi(q)$ é a função de Euler), enquanto que medidas pseudo-aleatórias devem estar equidistribuídas em todas as classes de congruências módulo q; como o quociente $\phi(q)/q$ pode ser feito arbitrariamente pequeno, vemos que não existem majorantes pseudo-aleatórios da função de van Mangoldt.

[3]Apesar de pedirmos crescimento lento para $w(n)$, pode-se constatar que, ao revisar os argumentos do capítulo 3, basta tomar $w(n)$ uma constante bem grande (dependendo apenas de k mas não de N).

Em outras palavras, $\widetilde{\Lambda}$ é uma medida. Nesta situação, o segundo resultado chave do trabalho de Green-Tao [5, Proposition 9.1] (baseado nos trabalhos de Goldston-Yıldırım) é:

Teorema 2.2.2. *Se $\epsilon_k = 1/(k+4)!2^k$ e N é um primo grande então existe ν uma medida k-pseudoaleatória tal que $\nu(n) \geq 2^{-k-5}k^{-1}\widetilde{\Lambda}(n)$ para $\epsilon_k N \leq n \leq 2\epsilon_k N$.*

Assim como boa parte dos resultados importantes sobre os números primos, a prova do teorema 2.2.2 passa pelo uso de certas regiões livre de zeros da função zeta de Riemann. Porém, para não interrompermos o fluxo de idéias, deixaremos para o capítulo 3 deste livro a demonstração completa do teorema 2.2.2.

Finalmente, assumindo momentaneamente os teoremas 2.2.1 e 2.2.2, demonstraremos o teorema de Green e Tao.

2.2.1 Prova do teorema de Green e Tao

Suponha que os teoremas 2.2.1 e 2.2.2 sejam válidos.

Se

$$f(n) = \frac{1}{k2^{k+5}}\widetilde{\Lambda}(n)1_{[\epsilon_k N, 2\epsilon_k N]}$$

então:

$$\mathbb{E}(f) = \frac{1}{Nk2^{k+5}} \sum_{\epsilon_k N \leq n \leq 2\epsilon_k N} \widetilde{\Lambda}(n) = \frac{1}{k2^{k+5}}\epsilon_k(1 + o(1)).$$

Como o teorema 2.2.2 garante a existência de uma medida k-pseudoaleatória majorando $2^{-k-5}k^{-1}\widetilde{\Lambda}$ em $[\epsilon_k N, 2\epsilon_k N]$, podemos aplicar o teorema 2.2.1 de Green-Tao-Szemerédi para concluir que:

$$\mathbb{E}(f(n)\dots f(n+(k-1)r)|n, r \in \mathbb{Z}_N) \geq c(k, k^{-1}2^{-k-3}\epsilon_k) - o(1).$$

Como o caso $r = 0$ contribui com um fator $O(\frac{1}{N}\log^k N) = o(1)$ na expressão, obtemos a existência de uma P.A. em $\mathbb{Z}_N$ (tomando N grande). Mais ainda, sendo $\epsilon_k < 1/k$ e $k \geq 3$ temos que essa P.A. é uma P.A. legítima de inteiros (e não apenas uma k-PA em $\mathbb{Z}_N$).

2.2.2 Alguns comentários

Uma vez que já reduzimos o teorema de Green e Tao aos teoremas 2.2.1 e 2.2.2, dedicaremos o resto deste capítulo a prova do teorema 2.2.1 de Green-Tao-Szemerédi (enquanto que a prova do teorema 2.2.2 ficará para o capítulo 3).

Entrentanto, para ilustrar as idéias por trás da prova do teorema de Green-Tao-Szemerédi (as quais podem ser técnicas e duras numa primeira leitura), faremos a prova do teorema de Roth (ou seja, do teorema de Szemerédi no caso $k = 3$). Em seguida, passaremos a demonstração propriamente dita do teorema de Green-Tao-Szemerédi e assim encerraremos o presente capítulo.

2.3 Prova do teorema de Roth

Conforme dissemos na introdução, para sentir o sabor da prova do teorema de Green-Tao-Szemerédi, vamos ver um pequeno argumento para encontrar 3-PA em conjuntos de densidade positiva, o qual, apesar de ser bem particular e específico, contém boa parte das idéias que iremos estudar em seguida.

Definimos $\Lambda_3(f,g,h) = \mathbb{E}(f(n)g(n+r)h(n+2r) : n, r \in \mathbb{Z}_N)$. O teorema de Roth pode ser reformulado assim:

Teorema 2.3.1. *Para toda $f : \mathbb{Z}_N \to \mathbb{R}$ não-negativa com*

$$0 < \delta \leq \|f\|_{L^1(\mathbb{Z}_N)} \leq \|f\|_{L^\infty(\mathbb{Z}_N)} \leq 1$$

tem-se

$$\Lambda_3(f,f,f) \geq c(3,\delta) - o_\delta(1).$$

Em outras palavras, queremos cotas inferiores para $\Lambda_3(f,f,f)$. Começamos com a observação simples de que cotas superiores são bem "fáceis" de se obter: por exemplo, pela desigualdade de Young,

$$|\Lambda_3(f,g,h)| \leq \|f\|_{L^p}\|g\|_{L^q}\|h\|_{L^r},$$

se $1 \leq p,q,r \leq \infty$ e $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} \leq 2$.

Por outro lado, estamos interessados apenas em cotas *inferiores* para Λ_3 e, *a priori*, as estimativas superiores não parecem ser úteis

para provar cotas inferiores. Entretanto, podemos decompor f como uma parte "boa" $g = \mathbb{E}(f)$ e outra parte "ruim" $b = f - \mathbb{E}(f)$. Usando a multilinearidade de Λ_3, podemos decompor $\Lambda_3(f,f,f)$ em oito pedaços:

$$\Lambda_3(f,f,f) = \Lambda_3(g,g,g) + \cdots + \Lambda_3(b,b,b).$$

Por hipótese, $\mathbb{E}(f) \geq \delta$, donde o primeiro termo é $\Lambda_3(g,g,g) \geq \delta^3$. Logo, boas cotas superiores dos outros termos (p.ex., a soma dos sete termos restantes são de magnitude menor que δ^3) nos levariam a concluir o teorema de Roth.

Porém, a estimativa acima (via desigualdade de Young) não é boa o suficiente, a menos que δ esteja próximo a 1 (digamos $\delta > 2/3$). Mas, podemos refinar nosso argumento de cotas superiores usando a *transformada de Fourier*:

$$\widehat{f}(\xi) := \mathbb{E}(f(x)e_N(-x\xi) : x \in \mathbb{Z}_N),$$

onde $e_N(x) := \exp(2\pi i x/N)$. Da fórmula de inversão

$$f(x) = \sum_{\xi \in \mathbb{Z}_N} \widehat{f}(\xi) e_N(x\xi)$$

obtemos que

$$\Lambda_3(f,g,h) = \sum_{\xi_1,\xi_2,\xi_3} \widehat{f}(\xi_1)\widehat{g}(\xi_2)\widehat{h}(\xi_3) \times \\ \mathbb{E}(e_N(n\xi_1 + (n+r)\xi_2 + (n+2r)\xi_3) : n, r \in \mathbb{Z}_N).$$

As esperanças no lado direito acima são 1 se $\xi_1 = \xi_3$ e $\xi_2 = -2\xi_1$ e 0 caso contrário. Em particular,

$$\Lambda_3(f,g,h) = \sum_{\xi \in \mathbb{Z}_N} \widehat{f}(\xi)\widehat{g}(-2\xi)\widehat{h}(\xi).$$

Da fórmula de Plancherel $\|f\|_{L^2(\mathbb{Z}_N)} = \|\widehat{f}\|_{l^2(\mathbb{Z}_N)}$ e da desigualdade de Hölder, segue que

$$|\Lambda_3(f,g,h)| \leq \|f\|_{L^2(\mathbb{Z}_N)} \|\widehat{g}\|_{l^4(\mathbb{Z}_N)} \|\widehat{h}\|_{l^4(\mathbb{Z}_N)}. \tag{2.3.1}$$

Usando esta estimativa, podemos provar que:

Proposição 2.3.1. *Seja f com uma decomposição $f = g + b$ onde*

$$\|g\|_{L^\infty(\mathbb{Z}_N)}, \|b\|_{L^\infty(\mathbb{Z}_N)} = O(1)$$

e

$$\|g\|_{L^1(\mathbb{Z}_N)}, \|b\|_{L^1(\mathbb{Z}_N)} = O(\delta).$$

Então

$$\Lambda_3(f,f,f) = \Lambda_3(g,g,g) + O(\delta^{5/4}\|\widehat{b}\|_{l^4(\mathbb{Z}_N)})$$

e

$$\Lambda_3(f,f,f) = \Lambda_3(g,g,g) + O(\delta\|\widehat{b}\|_{l^\infty(\mathbb{Z}_N)}).$$

Demonstração. Por hipótese, $\|g\|_{L^2(\mathbb{Z}_N)}, \|b\|_{L^2(\mathbb{Z}_N)} = O(\delta^{1/2})$, donde uma aplicação de Plancherel nos diz que

$$\|\widehat{g}\|_{l^2(\mathbb{Z}_N)}, \|\widehat{b}\|_{l^2(\mathbb{Z}_N)} = O(\delta^{1/2}).$$

Além disso, as cotas L^1 de g e b implicam

$$\|\widehat{g}\|_{l^\infty(\mathbb{Z}_N)}, \|\widehat{b}\|_{l^\infty(\mathbb{Z}_N)} = O(\delta).$$

Logo, a desigualdade de Hölder nos conduz a

$$\|\widehat{g}\|_{l^4(\mathbb{Z}_N)}, \|\widehat{b}\|_{l^4(\mathbb{Z}_N)} = O(\delta^{3/4}).$$

A proposição segue decompondo $\Lambda_3(f,f,f)$ em oito partes e usando (2.3.1). □

Esta proposição sugere que uma estratégia para conseguir cotas inferiores não-triviais de $\Lambda_3(f,f,f)$ passa por decompor $f = g+b$ em uma função boa g tendo o valor $\Lambda_3(g,g,g)$ "alto" e uma função ruim b com norma l^4 da sua transformada de Fourier pequena.

O leitor atento percebeu que já indicamos uma possibilidade de decomposição: $g = \mathbb{E}(f)$ e $b = f - \mathbb{E}(f)$. Note que temos a cota $\Lambda_3(g,g,g) \geq \delta^3$, uma estimativa relativamente boa, mas não temos boas cotas para $\|\widehat{b}\|_{l^4(\mathbb{Z}_N)}$; as nossas melhores cotas até o momento são $O(\delta^{3/4})$, o que é ruim pois permite que o erro domine o termo principal.[4]

[4]Com efeito, $f = \chi_{[1,\delta N]}$ tem $\Lambda_3(f,f,f) \sim \delta^2$ e $\Lambda_3(g,g,g) = \delta^3$, por exemplo.

Entretanto, podemos eliminar o caso de $b = f - \mathbb{E}(f)$ *linearmente uniforme*, i.e., $\|\widehat{b}\|_{l^\infty} \leq \delta^2/100$. O problema é o que fazer se b não é linearmente uniforme. Neste caso, adota-se a idéia de usar um *argumento de incremento na energia*, i.e.,

- **Argumento de incremento de energia**: uma vez que b não é uniforme, trocamos g por uma função com norma L^2 maior. Ao final de um número finito de passos, espera-se chegar a uma função b uniforme (já que a energia é finita).

Logicamente, esta idéia tem que ser trabalhada em detalhes para ver que ela conduz ao fim da prova do teorema de Roth. Para isto, vamos introduzir a definição:

Definição 2.3.1. *Dado K inteiro positivo, chamaremos as funções $f : \mathbb{Z}_N \to \mathbb{C}$ da forma*

$$f(x) = \sum_{j=1}^{K} c_j \exp(2\pi i \xi_j x/N),$$

onde $|c_j| \leq 1$ e $\xi_j \in \mathbb{Z}_N$ de funções K-quase-periódicas. *Além disso, fixado $\sigma > 0$, dizemos que uma função f é* (σ, K)-quase-periódica *se $\|f - f_{qp}\|_{L^2(\mathbb{Z}_N)} \leq \sigma$ para alguma função K-quase-periódica f_{qp}.*

Observação 2.3.1. *Note que se f e g são funções (σ, K)-quase-periódicas então fg é $(2\sigma, K^2)$-quase-periódica.*

O ponto importante no conceito de funções f quase-periódicas é que podemos obter boas cotas inferiores de $\Lambda_3(f, f, f)$ neste caso:

Lema 2.3.1 ("Múltipla Recorrência" de funções quase-periódicas)**.** *Dados $0 < \delta < 1$, $M \geq 1$, $0 < \sigma \leq \delta^3/100M$ e $0 \leq f \leq M$ uma função não-negativa limitada (σ, K)-quase-periódica com $\mathbb{E}(f) \geq \delta$, então*

$$\Lambda_3(f, f, f) \geq c(K, M, \delta) - o_{K,M,\delta}(1),$$

para algum $c(K, M, \delta) > 0$.

Demonstração. Seja $f_{qp}(x) = \sum_{j=1}^{K} c_j \exp(2\pi i x \xi_j/N)$ uma função K-quase-periódica aproximando f e tome $\varepsilon = \varepsilon(K, \delta) > 0$ uma constante pequena. Pelo teorema de aproximação simultânea de Dirichlet

(o qual decorre do príncipio da casa de pombos), temos

$$\mathbb{E}(\|r\xi_j\| \leq \varepsilon \text{ para todo } j\,;\, r \in \mathbb{Z}_N) \geq c(\varepsilon, K). \tag{2.3.2}$$

Aqui é fundamental ressaltar que a constante $c(\varepsilon, K) > 0$ *não depende* de N. Por outro lado, considerando a dinâmica $T(x) := x + 1$ e fixando r tal que $\|r\xi_j\| \leq \varepsilon$ (onde $1 \leq j \leq K$), temos:

$$\|f_{qp} \circ T^r - f_{qp}\|_{L^2(\mathbb{Z}_N)} \leq C(K)\varepsilon,$$

Isto combinado com a desigualdade triangular implica:

$$\|f \circ T^r - f\|_{L^2(\mathbb{Z}_N)} \leq \delta^3/50M + C(K)\varepsilon.$$

Aplicando T^r novamente na estimativa acima obtemos também

$$\|f \circ T^{2r} - f \circ T^r\|_{L^2(\mathbb{Z}_N)} \leq \delta^3/50M + C(K)\varepsilon.$$

Como a função f é limitada, destas estimativas decorre que:

$$\|f \cdot (f \circ T^r) \cdot (f \circ T^{2r}) - f^3\|_{L^1(\mathbb{Z}_N)} \leq \delta^3/2 + C(K)M\varepsilon.$$

Entretanto, sendo $f \geq 0$, usando nossa hipótese $\mathbb{E}(f) \geq \delta$ e a desigualdade de Hölder, obtemos

$$\|f^3\|_{L^1(\mathbb{Z}_N)} \geq \|f\|^3_{L^1(\mathbb{Z}_N)} \geq \delta^3.$$

Logo, $\mathbb{E}(f \cdot (f \circ T^r) \cdot (f \circ T^{2r})) \geq \delta^3/2 - C(K)M\varepsilon$. Escolhendo $\varepsilon > 0$ pequeno dependendo de δ, K e M, segue que

$$\mathbb{E}(f \cdot (f \circ T^r) \cdot (f \circ T^{2r})) \geq \delta^3/4.$$

Como $f \geq 0$, tomando a média em r e usando (2.3.2), podemos concluir

$$\mathbb{E}(f(n) \cdot (f \circ T^r)(n) \cdot (f \circ T^{2r})(n) \,|\, n, r \in \mathbb{Z}_N) \geq \delta^3 c(\varepsilon, K)/4.$$

Sendo $\Lambda_3(f, f, f) = \mathbb{E}(f(n) \cdot (f \circ T^r)(n) \cdot (f \circ T^{2r})(n) \,|\, n, r \in \mathbb{Z}_N)$, terminamos a demonstração do lema. □

Visando a utilização deste lema, estaremos interessados em decompor funções arbitrárias em uma soma de uma função quase-periódica e uma função linearmente uniforme. Com este intuito, veremos como construir sigma-álgebras cujas funções mensuráveis sejam todas quase-periódicas:

Lema 2.3.2. *Sejam $0 < \varepsilon \ll 1$ e χ uma função da forma $\chi(x) := \exp(2\pi i x\xi/N)$. Então, existe uma sigma-álgebra $\mathcal{B}_{\varepsilon,\chi}$ tal que $\|\chi - \mathbb{E}(\chi|\mathcal{B}_{\varepsilon,\chi})\|_{L^\infty(\mathbb{Z}_N)} \leq C\varepsilon$ e, para todo $\sigma > 0$ existe $K = K(\sigma,\varepsilon) > 0$ com a seguinte propriedade: toda f função $\mathcal{B}_{\varepsilon,\chi}$-mensurável satisfazendo a estimativa $\|f\|_{L^\infty(\mathbb{Z}_N)} \leq 1$ é (σ,K)-quase-periódica.*

Demonstração. Seguiremos um processo randômico para obter a sigma-álgebra desejada: tome α um número complexo no quadrado unitário e seja $\mathcal{B}_{\varepsilon,\chi}$ a sigma-álgebra cujos átomos tem a forma $\chi^{-1}(Q)$, onde Q é um quadrado tal que os vértices de $Q - \varepsilon\alpha$ estão sobre $\varepsilon\mathbb{Z}^2$. Note que esta sigma-álgebra possui $O(1/\varepsilon)$ átomos e $\|\chi - \mathbb{E}(\chi|\mathcal{B}_{\varepsilon,\chi})\|_{L^\infty(\mathbb{Z}_N)} \leq C\varepsilon$. Em particular, resta apenas verificar a segunda parte do lema para finalizar a prova. Observe que basta verificar o fato desejado para $\sigma = 2^{-n}$ (onde $n \gg 1$) com probabilidade $1 - O(\sigma)$ em α. Como $\mathcal{B}_{\varepsilon,\chi}$ possui $O(1/\varepsilon)$ átomos, é suficiente considerar o caso de f igual a função característica de um átomo A de $\mathcal{B}_{\varepsilon,\chi}$ e provar a propriedade desejada com probabilidade $1 - O(c(\varepsilon)\sigma)$. Note que nesta situação podemos reescrever f como $f(x) = 1_Q(\chi(x) - \varepsilon\alpha)$. Aplicando o teorema de aproximação de Weierstrass no disco $|z| \leq O(1/\varepsilon)$, encontramos um polinômio $P(z,\overline{z})$ com $C(\sigma,\varepsilon)$ termos e coeficientes limitados por $C(\sigma,\varepsilon)$ tal que $|P| \leq 1$ no disco $|z| \leq O(1/\varepsilon)$ e $|1_Q(z) - P(z,\overline{z})| = O(c(\varepsilon)\sigma)$ para todo z neste disco exceto por um conjunto de medida $O(c(\varepsilon)^2\sigma^2)$. Isto implica que

$$\|1_Q(\chi(x) - \varepsilon\alpha) - P(\chi(x) - \varepsilon\alpha, \overline{\chi(x)} - \varepsilon\overline{\alpha})\|_{L^2(\mathbb{Z}_N)} \leq c(\varepsilon)\sigma$$

com probabilidade $1 - O(c(\varepsilon)\sigma)$ em α. Porém $P(\chi(x) - \varepsilon\alpha, \overline{\chi(x)} - \varepsilon\overline{\alpha})$ é uma combinação linear de $C(\varepsilon,\sigma)$ funções da forma $\exp(\overline{2\pi i x\xi/N})$ com coeficientes limitados por $C(\varepsilon)$. Ou seja, $P(\chi(x) - \varepsilon\alpha, \overline{\chi(x)} - \varepsilon\overline{\alpha})$ é uma função K-quase-periódica (onde $K = C(\varepsilon,\sigma)$). Em particular, f é uma função (σ,K)-quase-periódica. Isto conclui a prova deste lema. □

Um corolário útil deste lema é:

Corolário 2.3.1. *Sejam $0 < \varepsilon_j \ll 1$ e $\chi_j(x) = \exp(2\pi i x \xi_j/N)$, onde $j = 1, \dots, n$. Denote por $\mathcal{B}_{\varepsilon_j,\chi_j}$ as sigma-álgebras fornecidas pelo lema acima. Então, para todo $\sigma > 0$, existe $K = K(n, \sigma, \varepsilon_1, \dots, \varepsilon_n)$ tal que toda f função $\mathcal{B}_{\varepsilon_1,\chi_1} \vee \cdots \vee \mathcal{B}_{\varepsilon_n,\chi_n}$-mensurável satisfazendo a estimativa $\|f\|_{L^\infty(\mathbb{Z}_N)} \leq 1$ é (σ, K)-quase-periódica.*

Demonstração. Como o número de átomos na sigma-álgebra $\mathcal{B}_{\varepsilon_1,\chi_1} \vee \cdots \vee \mathcal{B}_{\varepsilon_n,\chi_n}$ é $C(n, \varepsilon_1, \dots, \varepsilon_n)$, basta provar o corolário no caso em que f é a função característica de um átomo de $\mathcal{B}_{\varepsilon_1,\chi_1} \vee \cdots \vee \mathcal{B}_{\varepsilon_n,\chi_n}$. Porém, nesta situação f é o produto de n funções de características de átomos das sigma-álgebras $\mathcal{B}_{\varepsilon_j,\chi_j}$. Logo, o corolário segue diretamente da combinação do lema anterior com a observação 2.3.1. □

Outra propriedade interessante destas sigma-álgebras (além de conter funções quase-periódicas) é a identificação de *obstruções para a uniformidade linear*:

Lema 2.3.3. *Sejam b uma função limitada com $\|\widehat{b}\|_{l^\infty(\mathbb{Z}_N)} \geq \sigma > 0$ e $0 < \varepsilon \ll \sigma$. Então, existe uma função da forma $\chi(x) = \exp(2\pi i x \xi/N)$ tal que a sigma-álgebra $\mathcal{B}_{\varepsilon,\chi}$ associada satisfaz*

$$\|\mathbb{E}(b|\mathcal{B}_{\varepsilon,\chi})\|_{L^2(\mathbb{Z}_N)} \geq \sigma/2.$$

Demonstração. Por definição, existe uma frequência ξ tal que $|\widehat{b}(\xi)| \geq \sigma$, i.e.,

$$|\mathbb{E}(b(n) \exp(-2\pi i n \xi/N)|n \in \mathbb{Z}_N)| \geq \sigma.$$

Definimos $\chi(x) := \exp(2\pi i x \xi/N)$ e reescrevemos a desigualdade acima como

$$|\langle b, \chi \rangle_{L^2(\mathbb{Z}_N)}| \geq \sigma.$$

Pelo lema anterior, sabemos que existe uma sigma-álgebra $\mathcal{B}_{\varepsilon,\chi}$ tal que

$$\|\chi - \mathbb{E}(\chi|\mathcal{B}_{\varepsilon,\chi})\|_{L^\infty(\mathbb{Z}_N)} \leq C\varepsilon.$$

Por outro lado, sendo b limitado e a esperança condicional auto-adjunta, podemos combinar as duas estimativas acima para concluir que

$$\langle \mathbb{E}(b|\mathcal{B}_{\varepsilon,\chi}), \chi \rangle = \langle b, \mathbb{E}(\chi|\mathcal{B}_{\varepsilon,\chi}) \rangle \geq \sigma - C\varepsilon.$$

Isto implica que $\|\mathbb{E}(b|\mathcal{B}_{\varepsilon,\chi})\|_{L^2(\mathbb{Z}_N)} \geq \sigma - C\varepsilon \geq \sigma/2$, o que finaliza a prova do lema. □

O último ingrediente para a prova completa do teorema de Roth é a seguinte proposição de estrutura:

Proposição 2.3.2 ("teorema quantitativo de Koopman-von Neumann")**.** *Sejam $F : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ uma função qualquer, $0 < \delta \leq 1$, f uma função não-negativa limitada satisfazendo $\mathbb{E}(f) \geq \delta$ e $\sigma := \delta^3/100$. Então, existem uma constante $0 < K \leq C(\delta, F)$ e uma decomposição $f = g+b$ tais que g é limitada não-negativa, $\mathbb{E}(g) = \mathbb{E}(f)$, g é (σ, K)-quase-periódica e b verifica*

$$\|\widehat{b}\|_{l^\infty(\mathbb{Z}_N)} \leq F(\delta, K). \tag{2.3.3}$$

Demonstração. A idéia será utilizar o *argumento de incremento de energia* para construir g e b. Para esta construção, necessitaremos de duas sigma-álgebras $\mathcal{B}$ e $\widetilde{\mathcal{B}}$ as quais sempre terão a forma $\mathcal{B}_{\varepsilon_1,\chi_1} \vee \cdots \vee \mathcal{B}_{\varepsilon_n,\chi_n}$ durante todo o argumento. Mais ainda, iremos querer estimativas do tipo

$$\|\mathbb{E}(f|\widetilde{\mathcal{B}})\|^2_{L^2(\mathbb{Z}_N)} \leq \|\mathbb{E}(f|\mathcal{B})\|^2_{L^2(\mathbb{Z}_N)} + \sigma^2/4. \tag{2.3.4}$$

Observe que, pelo teorema de Pitágoras, a estimativa acima equivale a

$$\|\mathbb{E}(f|\widetilde{\mathcal{B}}) - \mathbb{E}(f|\mathcal{B})\|^2_{L^2(\mathbb{Z}_N)} \leq \sigma/2.$$

A prova desta proposição utilizará o seguinte algoritmo:

- *Estágio 0*: Começamos com $\mathcal{B}$ e $\widetilde{\mathcal{B}}$ iguais a sigma-álgebra trivial $\{0, \mathbb{Z}_N\}$. Note que a desigualdade (2.3.4) é satisfeita automaticamente neste estágio.

- *Estágio 1*: Considere $\mathcal{B}$ uma sigma-álgebra da forma $\mathcal{B}_{\varepsilon_1,\chi_1} \vee \cdots \vee \mathcal{B}_{\varepsilon_n,\chi_n}$, onde $\chi_j(x) = \exp(2\pi i x\xi_j/N)$. Sendo a função $\mathbb{E}(f|\mathcal{B})$ uma função limitada e $\mathcal{B}$-mensurável, o corolário 2.3.1 diz que podemos encontrar K dependendo de $\delta, n, \varepsilon_1, \ldots, \varepsilon_n$ tal que $\mathbb{E}(f|\mathcal{B})$ é $(\sigma/2, K)$-quase-periódica.

- *Estágio 2*: Fazemos $g = \mathbb{E}(f|\widetilde{\mathcal{B}})$ e $b = f - \mathbb{E}(f|\widetilde{\mathcal{B}})$. Se $\|\widehat{b}\|_{l^\infty(\mathbb{Z}_N)} \leq F(\delta, K)$, encerramos o algoritmo. Caso contrário, vamos para o estágio 3;

- *Estágio 3*: Como não terminamos o algoritmo no estágio 2, temos $\|\widehat{b}\|_{l^\infty} > F(\delta, K)$. Pelo lema 2.3.3, podemos encontrar $\varepsilon \ll F(\delta, K)$ e uma função χ da forma $\chi(x) = \exp(2\pi i x\xi/N)$ com uma sigma-álgebra associada $\mathcal{B}_{\varepsilon,\chi}$ tal que

$$\|\mathbb{E}(b|\mathcal{B}_{\varepsilon,\chi})\|_{L^2(\mathbb{Z}_N)} \geq F(\delta, K)/2.$$

  Da identidade

$$\mathbb{E}(b|\mathcal{B}_{\varepsilon,\chi}) = \mathbb{E}(\mathbb{E}(f|\widetilde{\mathcal{B}} \vee \mathcal{B}_{\varepsilon,\chi}) - \mathbb{E}(f|\widetilde{\mathcal{B}})|\mathcal{B}_{\varepsilon,\chi})$$

  e do teorema de Pitágoras tiramos também que

$$\|\mathbb{E}(f|\widetilde{\mathcal{B}} \vee \mathcal{B}_{\varepsilon,\chi}) - \mathbb{E}(f|\widetilde{\mathcal{B}})\|_{L^2(\mathbb{Z}_N)} \geq F(\delta, K)/2.$$

  Aplicando o teorema de Pitágoras novamente, obtemos a *estimativa de incremento de energia*:

$$\|\mathbb{E}(f|\widetilde{\mathcal{B}} \vee \mathcal{B}_{\varepsilon,\chi})\|^2_{L^2(\mathbb{Z}_N)} \geq \|\mathbb{E}(f|\widetilde{\mathcal{B}})\|^2_{L^2(\mathbb{Z}_N)} - F(\delta, K)^2/4.$$

- *Estágio 4*: Trocamos $\widetilde{\mathcal{B}}$ por $\widetilde{\mathcal{B}} \vee \mathcal{B}_{\varepsilon,\chi}$. Caso ainda tenhamos a estimativa (2.3.4), retornamos para o estágio 2; caso contrário, trocamos $\mathcal{B}$ por $\widetilde{\mathcal{B}}$ e vamos para o estágio 1.

Afirmo que este algoritmo termina. Com efeito, fixado $\mathcal{B}$ (e consequentemente K), cada vez que passamos pelo estágio 4 a *energia* $\|\mathbb{E}(f|\widetilde{\mathcal{B}})\|^2_{L^2(\mathbb{Z}_N)}$ aumenta de $F(\delta, K)^2/4$. Logo, o algoritmo termina ou a estimativa (2.3.4) é violada em $C(\delta, K, F) = C\sigma^2/F(\delta, K)^2$ passos. No segundo caso, trocamos $\mathcal{B}$ pela sigma-álgebra associada às $C(\delta, K, F)$ funções χ e parâmetros $\varepsilon \geq C(\delta, F, K)^{-1}$ aparecendo neste processo. Isto implica que a quantidade K associada a esta nova sigma-álgebra $\mathcal{B}$ será trocada por uma quantidade da forma $C(\delta, K, F)$ e a energia $\|\mathbb{E}(f|\mathcal{B})\|^2_{L^2(\mathbb{Z}_N)}$ cresceu pelo menos $\sigma^2/4$ graças a violação de (2.3.4). Por outro lado, o fato de f ser limitada garante que esta energia *não* pode ser maior que $O(1)$. Logo, estas trocas de $\mathcal{B}$ descritas acima só podem ser feitas no máximo $O(\sigma^{-2})$ vezes. Juntando estas informações vemos que o algoritmo inteiro termina em $C(\delta, F)$ passos (e a quantidade K nunca ultrapassa $C(\delta, F)$ durante todo o processo). Isto conclui a prova da proposição. □

Finalmente, vejamos como encerrar a demonstração do teorema de Roth: seja $F : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ uma função que escolheremos em alguns instantes e apliquemos a proposição acima para decompor $f = g + b$. Pelo lema 2.3.1, sabemos que

$$\Lambda_3(g, g, g) \geq c(\delta, K) - o_{\delta,K}(1).$$

Combinando esta desigualdade com (2.3.3) e a proposição 2.3.1, temos

$$\Lambda_3(f, f, f) \geq c(\delta, K) + O(\delta \cdot F(\delta, K)) - o_{\delta,K}(1).$$

Tomando F "suficientemente pequena", podemos absorver a segunda parcela do lado direito pela primeira parcela, de maneira que

$$\Lambda_3(g, g, g) \geq c(\delta, K)/2 - o_{\delta,K}(1).$$

Como $K \leq C(\delta, F) = C(\delta)$, o teorema de Roth está provado.

Encerrando esta seção, recapitularemos abaixo os dois passos principais da prova do teorema de Roth (o qual inspirará a demonstração do teorema de Green-Tao-Szemerédi):

- **primeiro passo**: definir uma classe de normas (ditas normas de Gowers $\|.\|_{U^{k-1}}$) para controlar a esperança de uma k-PA estar no suporte de f; note que, pela proposição 2.3.1, no caso $k = 3$, a norma l^4 da transformada de Fourier é um bom candidato;[5]

- **segundo passo**: generalizar o processo de incremento na energia acima discutido.

2.4 Demonstração do teorema de Green-Tao-Szemerédi

Nesta última seção do presente capítulo, provaremos ao longo de várias subseções os resultados que nos ajudarão a formalizar as idéias acima. Entretanto, como as demonstrações são técnicas, o leitor pode

[5]De fato, no caso $k = 3$, a norma de Gowers $\|.\|_{U^2}$ é a norma l^4 da transformada de Fourier; veja a observação 2.4.1 na próxima subseção.

se perder um pouco dos nossos objetivos finais. Por isso, daqui em diante, ao final de cada subseção, faremos um "resumo" dos resultados provados e como eles se encaixam na estratégia de incremento de energia acima traçada.

2.4.1 Normas de Gowers

Seja $\{0,1\}^d$ o cubo discreto d-dimensional, e $w = (w_1, \ldots, w_d) \in \{0,1\}^d$. Se $h \in \mathbb{Z}_N^d$ então $w.h := w_1.h_1 + \ldots w_d.h_d$. Se $\{f_w\}_{w\in\{0,1\}^d}$ o *produto interno de Gowers* é:

$$\langle (f_w) \rangle_{U^d} := \mathbb{E}(\Pi_w f_w(n + w.h) | n \in \mathbb{Z}_N, h \in \mathbb{Z}_N^d).$$

A primeira observação é que se $f_w = f$ para todo w então $\langle (f_w) \rangle_{U^d} \geq 0$. Assim podemos definir as *normas de Gowers* (usando $f_w = f$):

$$\|f\|_{U^d} := \langle (f) \rangle_{U^d}^{\frac{1}{2^d}}.$$

Uma ferramenta basilar para a análise das normas de Gowers é a *desigualdade de Gowers-Cauchy-Schwarz*:

$$|\langle (f_w) \rangle_{U^d}| \leq \Pi_w \|f_w\|_{U^d}.$$

A prova desta desigualdade segue do fato de que, quando f_w não depende de w_d, vale a igualdade

$$\langle (f_w) \rangle_{U^d} = \mathbb{E}(\mathbb{E}(\prod_{w'\in\{0,1\}^{d-1}} f_{w',0}(y + w'.h') : y \in \mathbb{Z}_N) \times$$
$$\mathbb{E}(\prod_{w'\in\{0,1\}^{d-1}} f_{w',1}(y + w'.h') : y \in \mathbb{Z}_N | h' \in (\mathbb{Z}_N)^{d-1})).$$

Logo, por Cauchy-Schwarz, temos

$$|\langle (f_w) \rangle_{U^d}| \leq \langle (f_{w',0}) \rangle_{U^d}^{1/2} \langle (f_{w',1}) \rangle_{U^d}^{1/2}.$$

Como podemos trocar w_d por qualquer outro dígito, aplicando a desigualdade acima d vezes, obtemos a desigualdade de Gowers-Cauchy-Schwarz.

Além disso, a fórmula binomial e a multilinearidade do produto interno nos levam a *desigualdade triangular de Gowers*:

$$\|f+g\|_{U^d} \leq \|f\|_{U^d} + \|g\|_{U^d}.$$

Finalmente, temos a relação de monotonicidade:

$$\|f\|_{U^{d-1}} \leq \|f\|_{U^d},$$

a qual é uma consequência de Gowers-Cauchy-Schwarz aplicado a $f_w := 1$ quando $w_d = 1$ e $f_w := f$ quando $w_d = 0$.

Observação 2.4.1. *Como $\|.\|_{U^d}$ são homogêneas, acabamos de mostrar que $\|.\|_{U^d}$ são semi-normas. Entretanto, $\|.\|_{U1}$ não é norma pois $\|f\|_{U1} = \mathbb{E}(f)$. Porém, pode-se provar (por cálculo direto) que:*

$$\|f\|_{U^2} = (\sum \widehat{f}(\xi)^4)^{\frac{1}{4}},$$

onde $\widehat{f}(\xi) = \mathbb{E}(f(x)e^{-2\pi i\xi/N}; x \in \mathbb{Z}_N)$ e vale a fórmula de inversão $f(x) = \sum \widehat{f}(\xi)e^{2\pi ix\xi/N}$. Consequentemente, as normas de Gowers para $d \geq 2$ são normas genuínas.

Com esta notação, a generalização natural da proposição 2.3.1 é:

Teorema 2.4.1 (von Neumann generalizado)**.** *Se ν é uma medida k-pseudoaleatória e $f_0, \ldots, f_{k-1} \in L^1(\mathbb{Z}_N)$ são tais que $|f_j(x)| \leq 1 + \nu(x)$ então se $c_0, \ldots, c_{k-1} \in \mathbb{Z}_N$ são distintos, temos que:*

$$\mathbb{E}(\Pi_j f_j(n + c_j r)|n, r \in \mathbb{Z}_N) = O(\inf \|f_j\|_{U^{k-1}}) + o(1).$$

Demonstração. Começemos com algumas simplificações: a menos de trocar ν por $(\nu+1)/2$, rearranjar f_j, c_j e transladar x por $c_0 r$, podemos assumir que

$$|f_j(x)| \leq \nu(x),\ \forall\ x \in \mathbb{Z}_N, j = 0, \ldots, k-1,$$

$$\inf_{0 \leq j \leq k-1} \|f_j\|_{U^{k-1}} = \|f_0\|_{U^{k-1}}$$

e

$$c_0 = 0.$$

Isto reduz o problema a provar que

$$\mathbb{E}\left(\prod_{j=0}^{k-1} f_j(x+c_j r)|x,r\in\mathbb{Z}_N\right) = O(\|f_0\|_{U^{k-1}}) + o(1).$$

Dividiremos a demonstração desta igualdade em duas partes: na primeira provaremos uma desigualdade de Cauchy-Schwarz e a aplicaremos $k-1$ vezes ao lado esquerdo da igualdade acima, obtendo assim uma estimativa de $\mathbb{E}\left(\prod_{j=0}^{k-1} f_j(x+c_j r)|x,r\in\mathbb{Z}_N\right)$ por uma *soma com pesos* de f_0 sobre cubos $(k-1)$-dimensionais; na segunda mostraremos que a condição de formas lineares implica que estes pesos são iguais a 1 em média, o que nos permitirá que deduzir o resultado desejado.

Para enunciar a desigualdade de Cauchy-Schwarz de modo razoável, introduziremos um pouco de notação. Dados $0 \le d \le k-1$, dois vetores $y = (y_1,\dots,y_{k-1}) \in (\mathbb{Z}_N)^{k-1}$ e $y = (y'_{k-d},\dots,y'_{k-1}) \in (\mathbb{Z}_N)^d$, e um conjunto $S \subset \{k-d,\dots,k-1\}$, definimos o vetor $y^{(S)} = (y_1^{(S)},\dots,y_{k-1}^{(S)}) \in (\mathbb{Z}_N)^{k-1}$ por

$$y_i^{(S)} := \begin{cases} y_i & \text{se } i \notin S \\ y'_i & \text{se } i \in S. \end{cases}$$

Em outras palavras, S indica quais componentes de $y^{(S)}$ provém de y' ao invés de y.

Lema 2.4.1. *Sejam $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ uma medida e $\phi_0,\dots,\phi_{k-1} : (\mathbb{Z}_N)^{k-1} \to \mathbb{Z}_N$ funções das $k-1$ variáveis y_i tais que ϕ_i não depende de y_i para $1 \le i \le k-1$. Suponha que $f_0,\dots,f_{k-1} \in L^1(\mathbb{Z}_N)$ são funções com $|f_i(x)| \le \nu(x)$ para todo $x \in \mathbb{Z}_N$ e $0 \le i \le k-1$. Para cada $0 \le d \le k-1$ e $1 \le i \le k-1$, defina*

$$J_d := \mathbb{E}\left(\prod_{S\subset\{k-d,\dots,k-1\}} \Big(\prod_{i=0}^{k-d-1} f_i(\phi_i(y^{(S)}))\times \right.$$
$$\left. \times \prod_{i=k-d}^{k-1} \nu^{1/2}(\phi_i(y^{(S)}))\Big| y\in(\mathbb{Z}_N)^{k-1}, y'\in(\mathbb{Z}_N)^d\right),$$

e

$$P_d := \mathbb{E}\left(\prod_{S\subset\{k-d,\dots,k-1\}} \nu(\phi_{k-d-1}(y^{(S)}))|y\in(\mathbb{Z}_N)^{k-1}, y'\in(\mathbb{Z}_N)^d\right).$$

Então para todo $0\le d\le k-2$*, temos a desigualdade*

$$|J_d|^2 \le P_d J_{d+1}.$$

Prova do lema 2.4.1. Considere J_d. Como ϕ_{k-d-1} não depende de y_{k-d-1}, podemos tirar as quantidades dependentes de ϕ_{k-d-1} da média em y_{k-d-1}, o que nos permite escrever

$$J_d = \mathbb{E}(G(y,y')H(y,y')|y_1,\dots,y_{k-d-2},y_{k-d},\dots,y_{k-1}, y'_{k-d},\dots,y'_{k-1}\in\mathbb{Z}_N),$$

onde

$$G(y,y') := \prod_{S\subset\{k-d,\dots,k-1\}} f_{k-d-1}(\phi_{k-d-1}(y^{(S)}))\nu^{-1/2}(\phi_{k-d-1}(y^{(S)}))$$

e

$$H(y,y') := \mathbb{E}(\prod_{S\subset\{k-d,\dots,k-1\}}\prod_{i=0}^{k-d-2} f_i(\phi_i(y^{(S)})) \times \prod_{i=k-d-1}^{k-1}\nu^{1/2}(\phi_i(y^{(S)}))|y_{k-d-1}\in\mathbb{Z}_N).$$

Usando Cauchy-Schwarz,

$$|J_d|^2 \le \mathbb{E}(|G(y,y')|^2|y_1,\dots,y_{k-d-2},y_{k-d},\dots,y_{k-1}, y'_{k-d},\dots,y'_{k-1}\in\mathbb{Z}_N)\times \times \mathbb{E}(|H(y,y')|^2|y_1,\dots,y_{k-d-2},y_{k-d},\dots,y_{k-1}, y'_{k-d},\dots,y'_{k-1}\in\mathbb{Z}_N).$$

Por outro lado, como $|f_{k-d-1}(x)|\le\nu(x)$ para todo x,

$$\mathbb{E}(|G(y,y')|^2|y_1,\dots,y_{k-d-2},y_{k-d},\dots,y_{k-1},y'_{k-d},\dots,y'_{k-1}\in\mathbb{Z}_N)\le P_d.$$

Mais ainda, escrevendo a definição de $H(y, y')$ e expandindo os quadrados trocando a variável y_{k-d-1} pelas novas variáveis y_{k-d-1}, y'_{k-d-1}, vemos que

$$\mathbb{E}(|H(y,y')|^2 | y_1, \ldots, y_{k-d-2}, y_{k-d}, \ldots, y_{k-1}, y'_{k-d}, \ldots, y'_{k-1} \in \mathbb{Z}_N) = J_{d+1}.$$

Isto completa a prova. □

Aplicando este lema $(k-1)$ vezes, obtemos

$$|J_0|^{2^{k-1}} \leq J_{k-1} \prod_{d=0}^{k-2} P_d^{2^{k-2-d}}.$$

Observe que, por definição,

$$J_0 = \mathbb{E}\left(\prod_{i=0}^{k-1} f_i(\phi_i(y)) | y \in (\mathbb{Z}_N)^{k-1} \right).$$

Para provar a desigualdade desejada, escolhemos[6]

$$\phi_i(y) := \sum_{j=1}^{k-1} \left(1 - \frac{c_i}{c_j}\right) y_j$$

de modo que $\phi_0(y) = y_1 + \cdots + y_{k-1}$, $\phi_i(y)$ não dependem de y_i e, para todo y, $\phi(y) = x + c_i r$ onde

$$r = -\sum_{i=1}^{k-1} \frac{y_i}{c_i}.$$

Agora a transformação sobrejetiva $\Phi : (\mathbb{Z}_N)^{k-1} \to (\mathbb{Z}_N)^2$ definida por

$$\Phi(y) := (y_1 + \cdots + y_{k-1}, \frac{y_1}{c_1} + \cdots + \frac{y_{k-1}}{c_{k-1}})$$

[6]Estamos utilizando aqui que os números c_j são distintos.

tem número constante de pré-imagens, donde uma conta simples mostra que

$$\mathbb{E}\left(\prod_{j=0}^{k-1} f_j(x+c_j r)|x,r\in\mathbb{Z}_N\right)=\mathbb{E}\left(\prod_{i=0}^{k-1} f_i(\phi_i(y))|y\in(\mathbb{Z}_N)^{k-1}\right)=J_0.$$

Entretanto, $P_d=1+o(1)$ para cada $0\leq d\leq k-2$, pois ν satisfaz a $(2^d,k-1+d,k)$-condição de formas lineares. Em particular, das estimativas anteriores obtemos

$$J_0^{2^{k-1}}\leq(1+o(1))J_{k-1}.$$

Fixe y. Quando S varia sobre os subconjuntos de $\{1,\ldots,k-1\}$, $\phi_0(y^{(S)})$ varia no cubo $(k-1)$-dimensional $\{x+w\cdot h:w\in\{0,1\}^{k-1}\}$, onde $x=y_1+\cdots+y_{k-1}$ e $h_i=y_i'-y_i$, $i=1,\ldots,k-1$. Logo,

$$J_{k-1}=\mathbb{E}\left(W(x,h)\prod_{w\in\{0,1\}^{k-1}} f_0(x+w\cdot h)|x\in\mathbb{Z}_N,h\in(\mathbb{Z}_N)^{k-1}\right),$$

com o peso $W(x,h)$ dado por

$$\begin{aligned}W(x,h)=\mathbb{E}(&\prod_{w\in\{0,1\}}\prod_{i=1}^{k-1}\prod^{k-2}\nu^{1/2}(\phi_i(y+wh))\times\\&\nu^{1/2}(\phi_{k-1}(y+wh))|y_1,\ldots,y_{k-2}\in\mathbb{Z}_N)\\=\mathbb{E}(&\prod_{i=1}^{k-2}\prod_{w\in\{0,1\}^{k-1},w_i=0}\nu(\phi_i(y+wh))\times\\&\prod_{w\in\{0,1\}^{k-1},w_{k-1}=0}\nu(\phi_{k-1}(y+wh))|y_1,\ldots,y_{k-2}\in\mathbb{Z}_N),\end{aligned}$$

onde $wh\in(\mathbb{Z}_N)^{k-1}$ é o vetor com coordenadas $(wh)_j:=w_jh_j$ e $y\in(\mathbb{Z}_N)^{k-1}$ é o vetor com componentes y_j para $1\leq k-2$ e $y_{k-1}:=x-y_1-\cdots-y_{k-2}$. Porém, a definição da norma de Gowers dizem que

$$\mathbb{E}\left(\prod_{w\in\{0,1\}^{k-1}} f_0(x+w\cdot h)|x\in\mathbb{Z}_N,h\in(\mathbb{Z}_N)^{k-1}\right)=\|f_0\|_{U^{k-1}}^{2^{k-1}}.$$

Portanto, basta provar que

$$\mathbb{E}\left((W(x,h)-1)\prod_{w\in\{0,1\}^{k-1}} f_0(x+w\cdot h)|x\in\mathbb{Z}_N, h\in(\mathbb{Z}_N)^{k-1}\right)=o(1).$$

Como $|f_j(x)|\leq\nu(x)$, segue que é suficiente mostrar

$$\mathbb{E}\left(|W(x,h)-1|\prod_{w\in\{0,1\}^{k-1}} \nu(x+w\cdot h)|x\in\mathbb{Z}_N, h\in(\mathbb{Z}_N)^{k-1}\right)=o(1).$$

Por Cauchy-Schwarz, isto decorre imediatamente do seguinte lema:

Lema 2.4.2 (ν cobre uniformemente seus cubos)**.** *Para $n=0,2$, vale*

$$\mathbb{E}\left(|W(x,h)-1|^n\prod_{w\in\{0,1\}^{k-1}} \nu(x+w\cdot h)|x\in\mathbb{Z}_N, h\in(\mathbb{Z}_N)^{k-1}\right)=o(1).$$

Demonstração. Expandindo o quadrado, vemos que basta provar que, para $q=0,1,2$, vale

$$\mathbb{E}\left(W(x,h)^q\prod_{w\in\{0,1\}^{k-1}} \nu(x+w\cdot h)|x\in\mathbb{Z}_N, h\in(\mathbb{Z}_N)^{k-1}\right)=o(1).$$

Porém, isto é uma consequência da condição de formas lineares:

- no caso $q=0$, aplique a $(2^{k-1},k,1)$-condição de formas lineares com variáveis $x,h_1,\ldots,h_{k-1}$ e formas lineares $x+w\cdot h$, $w\in\{0,1\}^{k-1}$;
- no caso $q=1$, aplique a $(2^{k-2}(k+1),2k-2,k)$-condição de formas lineares com variáveis $x,h_1,\ldots,h_{k-1},y_1,\ldots,y_{k-2}$ e formas lineares

$$\begin{cases}\phi_i(y+w\cdot h), & w\in\{0,1\}^{k-1}, w_i=0 \text{ para } 1\leq i\leq k-1\\ x+w\cdot h, & w\in\{0,1\}^{k-1};\end{cases}$$

- no caso $q = 2$, aplique a $(k2^{k-1}, 3k-4, k)$-condição de formas lineares com variáveis

$$x, h_1, \ldots, h_{k-1}, y_1, \ldots, y_{k-2}, y'_1, \ldots, y'_{k-2}$$

e formas lineares

$$\begin{cases} \phi_i(y + w \cdot h), & w \in \{0,1\}^{k-1}, w_i = 0 \text{ para } 1 \leq i \leq k-1 \\ \phi_i(y' + w \cdot h), & w \in \{0,1\}^{k-1}, w_i = 0 \text{ para } 1 \leq i \leq k-1 \\ x + w \cdot h, & w \in \{0,1\}^{k-1}; \end{cases}$$

Aqui estamos adotando as convenções $y_{k-1} = x - y_1 - \cdots - y_{k-2}$ e $y'_{k-1} = x - y'_1 - \cdots - y'_{k-2}$. Claramente isto completa a prova do lema. □

Como dissemos antes, isto encerra a prova do teorema 2.4.1 de von Neumann generalizado. □

Observação 2.4.2. *Note que só utilizamos a condição de formas lineares na prova do teorema 2.4.1*

Encerrando o estudo das normas de Gowers desta subseção, enunciaremos um lema simples e útil sobre a distância de Gowers $\|.\|_{U^{k-1}}$ entre as medidas k-pseudo-aleatórias ν e $\nu_{const} \equiv 1$:

Lema 2.4.3. *Suponha que ν é uma medida k-pseudo-aleatória. Então,*

$$\|\nu - 1\|_{U^d} = o(1),$$

para todo $d \leq k-1$.

Demonstração. Observe que a condição de formas lineares para ν implicam facilmente que $\|\nu\|_{U^{k-1}} = 1 + o(1)$. Entretanto, podemos refinar um pouco mais este raciocínio. Com efeito, note que, pela monotonicidade das normas de Gowers, basta ver que $\|\nu - 1\|_{U^{k-1}} = o(1)$. Multiplicando por 2^{k-1}, reduzindo nosso problema a provar que

$$\mathbb{E}\left(\prod_{w \in \{0,1\}^{k-1}} \nu(x + w \cdot h) \middle| x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1} \right) = o(1).$$

O lado esquerdo da identidade acima pode ser expandido como

$$\sum_{A\subset\{0,1\}^{k-1}}(-1)^{|A|}\mathbb{E}\left(\prod_{w\in A}\nu(x+w\cdot h)\big|x\in\mathbb{Z}_N,h\in\mathbb{Z}_N^{k-1}\right).$$

Olhando para a expressão

$$\mathbb{E}\left(\prod_{w\in A}\nu(x+w\cdot h)\big|x\in\mathbb{Z}_N,h\in\mathbb{Z}_N^{k-1}\right)$$

para um $A\subset\{0,1\}^{k-1}$ fixado, vemos que ela toma a forma

$$\mathbb{E}\left(\nu(\phi_1(\mathbf{x}))\dots\nu(\phi_{|A|}(\mathbf{x}))\big|\mathbf{x}\in\mathbb{Z}_N^k\right),$$

onde $\mathbf{x}:=(x,h_1,\dots,h_{k-1})$ e $\phi_1,\dots,\phi_{|A|}$ são uma ordenação das $|A|$ formas lineares $x\mapsto x+w\cdot h$ com $w\in A$. Obviamente estas formas lineares não são múltiplas racionais entre si, donde a $(2^{k-1},k,1)$-condição de formas lineares pode ser aplicada para concluir a prova do lema. □

Façamos agora um resumo da discussão acima.

Resumo da subseção "Normas de Gowers":

Nesta subseção identificamos normas naturalmente associadas ao problema de contar progressões cujos elementos pertencem ao suporte de uma família dada de funções, a saber as normas de Gowers, e provamos o teorema 2.4.1, o qual diz que as normas de Gowers majoram o número de progressões no suporte de uma sequência de funções, a menos de um erro negligenciável. Como vimos no caso do teorema de Roth, esta majoração é importante para obter boas cotas por baixo, o nosso objetivo inicial. O próximo estágio será a introdução do conceito de anti-uniformidade, o qual desempenhará papel importante no momento de decompor nossas funções nas partes boa e ruim.

2.4.2 Anti-Uniformidade

Como as normas de Gowers para $d\geq 2$ são normas genuínas podemos tomar as normas duais:

$$\|g\|_{(U^{k-1})^*}:=\sup_{\|f\|_{U^{k-1}}\leq 1}|\langle f,g\rangle|,$$

onde $\langle f, g\rangle$ denota o produto L^2. Dizemos que g é *anti-uniforme* se $\|g\|_{(U^{k-1})^*} = O(1)$ e $\|g\|_{L^\infty} = O(1)$.

Observação 2.4.3. *Apesar de não pretendermos utilizar, note que no caso $k = 3$, a observação 2.4.1 nos dá a fórmula:*

$$\|g\|_{(U^2)^*} = \left(\sum_{\xi \in \mathbb{Z}_N} |\widehat{g}(\xi)|^{4/3} \right)^{3/4}.$$

Observe que se g é anti-uniforme e $|\langle f, g\rangle|$ é grande então f não pode ser uniforme pois $|\langle f, g\rangle| \leq \|f\|_{U^{k-1}} \|g\|_{(U^{k-1})^*}$. Logo temos uma obstrução para uniformidade.

Além disso temos uma maneira canônica de construir funções anti-uniformes: dada $F \in L^1(\mathbb{Z}_N)$, definimos a função dual de F como:

$$DF(x) := \mathbb{E}(\Pi_{w \neq 0} F(x + w.h) | h \in \mathbb{Z}_N^{k-1}).$$

Dentre as várias propriedades elementares destas funções, podemos citar:

Lema 2.4.4. *Seja ν uma medida k-pseudo-aleatória e $F \in L^1(\mathbb{Z}_\mathbb{N})$ uma função qualquer. Tem-se:*

- $\langle F, DF\rangle = \|F\|_{U^{k-1}}^{2^{k-1}}$,
- $\|DF\|_{(U^{k-1})^*} = \|F\|_{U^{k-1}}^{2^{k-1}-1}$ *e*
- *se $|F| \leq 1 + \nu$, então $\|DF\|_{L^\infty} \leq 2^{2^{k-1}-1} + o(1)$.*

Demonstração. A identidade $\langle F, DF\rangle = \|F\|_{U^{k-1}}^{2^{k-1}}$ segue diretamente das definições da norma de Gowers e DF, e deixamos como exercício para o leitor. Para provar a segunda identidade, considere $F \neq 0$ (já que o caso $F = 0$ é trivial) e note que a definição das normas duais combinadas com a identidade $\langle F, DF\rangle = \|F\|_{U^{k-1}}^{2^{k-1}}$ dizem que basta provar que uma função f qualquer vale

$$|\langle f, DF\rangle| \leq \|f\|_{U^{k-1}} \|F\|_{U^{k-1}}^{2^{k-1}-1}.$$

Porém, a definição de DF mostra que $\langle f, DF\rangle$ é o produto interno de Gowers $\langle (f_w)_{w \in \{0,1\}^{k-1}} \rangle_{U^{k-1}}$ onde $f_w := f$ quando $w = 0$ e $f_w := F$

caso contrário, donde a desigualdade acima segue da desigualdade de Gowers-Cauchy-Schwarz.

Finalmente, o último item segue da condição de formas lineares. De fato, como F é limitada por $2(1+\nu)/2 := 2\nu_{1/2}$, vemos que basta provar

$$D\nu_{1/2}(x) \leq 1 + o(1)$$

uniformemente para todo $x \in \mathbb{Z}_N$. Por outro lado, a definição de função dual diz que $D\nu_{1/2}$ pode ser expandido como

$$\mathbb{E}\left(\prod_{w \in \{0,1\}^{k-1} - \{0\}} \nu_{1/2}(x + w \cdot h) \,\middle|\, h \in \mathbb{Z}_N^{k-1} \right).$$

Como $\nu_{1/2}$ é uma medida k-pseudo-aleatória, segue da condição de formas lineares que este termo é $1 + o(1)$. □

Observação 2.4.4. *Este é o único ponto onde a condição de formas lineares com termos não-homogêneos $b_i \neq 0$ é aplicada; com efeito, na demonstração acima, todos os b_i são iguais a x.*

Chamaremos as funções DF, onde F é limitada (pontualmente) por $1 + \nu$, de *funções anti-uniformes básicas*; uma propriedade relevante destas funções é sua boa distribuição com respeito a ν:

Proposição 2.4.1. *Se ν é k-pseudoaletória, $\Phi : I^K \to \mathbb{R}$ é contínua e $DF_1, \ldots, DF_K$ funções anti-uniformes básicas, defina*

$$\Psi(x) = \Phi(DF_1(x), \ldots, DF_K(x)).$$

Então, $\langle \nu - 1, \Psi \rangle = o_{k,\Phi}(1)$. Além disso, a quantidade da direita pode ser tomada uniforme sobre um conjunto compacto de Φ's.

Demonstração. A idéia será usar o teorema de aproximação de Weierstrass e o fato de ν ser uma medida para reduzir nosso problema a provar esta proposição no caso mais "simples" de Φ ser um polinômio.

Como de costume, podemos trocar ν por $(\nu + 1)/2$ de modo que $|F_j(x)| \leq \nu(x)$ para todo $x \in \mathbb{Z}_N$, $1 \leq j \leq K$.

Lema 2.4.5. *Seja $d \geq 1$. Para todo polinômio P de grau d com coeficientes reais (independentes de N) vale*

$$\|P(DF_1, \ldots, DF_K)\|_{(U^{k-1})^*} = O_{K,d,P}(1).$$

Demonstração. Por linearidade e aumentando K para dK se necessário, basta provar o resultado para $P(x_1, \ldots, x_K) = x_1 \ldots x_K$. Ou seja, queremos ver que

$$\langle f, \prod_{j=1}^{K} DF_j \rangle = O_K(1)$$

para todo $f : \mathbb{Z}_N \to \mathbb{R}$ satisfazendo $\|f\|_{U^{k-1}} \leq 1$. Expandimos o lado esquerdo como

$$\mathbb{E}\left(f(x) \prod_{j=1}^{K} \mathbb{E}(\prod_{w \in \{0,1\}^{k-1} : w \neq 0} F_j(x + w \cdot h^{(j)}) | h^{(j)} \in (\mathbb{Z}_N)^{k-1}) \middle| x \in \mathbb{Z}_N \right)$$

Fazendo a mudança $h^{(j)} = h + H^{(j)}$ para todo $h \in (\mathbb{Z}_N)^{k-1}$, tomando a média em h, expandindo os produtos em j e intercambiando as esperanças, podemos reescrever isso em termos do produto interno de Gowers

$$\mathbb{E}(\langle (f_{w,H})_{w \in \{0,1\}^{k-1}} \rangle_{U^{k-1}} | H \in ((\mathbb{Z}_N)^{k-1})^K),$$

onde $H = (H^{(1)}, \ldots, H^{(K)})$, $f_{0,H} := f$, $f_{w,H} := g_{w \cdot H}$ para $w \neq 0$ com $w \cdot H = (w \cdot H^{(1)}, \ldots, w \cdot H^{(K)})$ e

$$g_{u^{(1)}, \ldots, u^{(K)}}(x) := \prod_{j=1}^{K} F_j(x + u^{(j)}) \text{ para todo } u^{(1)}, \ldots, u^{(K)} \in \mathbb{Z}_N.$$

Em particular, a desigualdade de Gowers-Cauchy-Schwarz e $\|f\|_{U^{k-1}} \leq 1$ reduzem o lema à prova da estimativa

$$\mathbb{E}\left(\prod_{w \in \{0,1\}^{k-1} : w \neq 0} \|g_{w \cdot H}\|_{U^{k-1}} | H \in ((\mathbb{Z}_N)^{k-1})^K \right) = O_K(1).$$

Por Hölder, basta ver que

$$\mathbb{E}(\|g_{w\cdot H}\|_{U^{k-1}}^{2^{k-1}-1}|H \in ((\mathbb{Z}_N)^{k-1})^K)) = O_K(1),$$

para cada $w \in \{0,1\}^{k-1} - 0$.

Fixe w. Como $2^{k-1} - 1 \leq 2^{k-1}$ e estamos em espaços de probabilidade, basta provar

$$\mathbb{E}(\|g_{w\cdot H}\|_{U^{k-1}}^{2^{k-1}}|H \in ((\mathbb{Z}_N)^{k-1})^K)) = O_K(1).$$

Esta última estimativa é verdadeira por um argumento assim: $w \neq 0$ implica que a transformação $w \to w \cdot H$ é recobrimento; isto permite usá-la para mudar as variáveis de maneira que o lado esquerdo da identidade acima é

$$\mathbb{E}(\|g_{u^{(1)},\ldots,u^{(K)}}\|_{U^{k-1}}^{2^{k-1}}|u^{(1)},\ldots,u^{(K)} \in \mathbb{Z}_N).$$

Usando as definições de norma de Gowers e $g_{u^{(1)},\ldots,u^{(K)}}$, podemos expandir este termo como

$$\mathbb{E}\left(\prod_{\widetilde{w}\in\{0,1\}^{k-1}}\prod_{j=1}^{K} F_j(x+u^{(j)}+h\cdot\widetilde{w}) \,\middle|\, x,u^{(1)},\ldots,u^{(K)} \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1}\right).$$

Fatorando, chegamos na expressão

$$\mathbb{E}\left(\prod_{j=1}^{K}\mathbb{E}(F_j(x+u^{(j)}+h\cdot\widetilde{w})\,|\,u^{(j)} \in \mathbb{Z}_N) \,\middle|\, x \in \mathbb{Z}_N,\, h \in \mathbb{Z}_N^{k-1}\right).$$

Usando a hipótese $|F_j(x)| \leq \nu(x)$, nossa tarefa fica reduzida a estimar

$$\mathbb{E}\left(\mathbb{E}(\nu(x+u+h\cdot\widetilde{w})\,|\,u \in \mathbb{Z}_N)^K \,\middle|\, x \in \mathbb{Z}_N,\, h \in \mathbb{Z}_N^{k-1}\right).$$

Fazendo a mudança de variáveis $y = x + u$ e tomando a média sobre x, nosso objetivo é provar que

$$\mathbb{E}\left(\mathbb{E}(\nu(y+h\cdot\widetilde{w})\,|\,y \in \mathbb{Z}_N)^K \,\middle|\, h \in \mathbb{Z}_N^{k-1}\right) = O_K(1).$$

Neste ponto, estamos prontos para usar a condição de correlação, a qual nos diz que

$$\mathbb{E}\left(\nu(y+h\cdot\widetilde{w})\,\middle|\, y\in\mathbb{Z}_N\right)\leq \sum_{\widetilde{w},\widetilde{w}'\in\{0,1\}^{k-1},\widetilde{w}\neq\widetilde{w}'}\tau(h\cdot(\widetilde{w}-\widetilde{w}')),$$

onde τ é uma função peso satisfazendo $\mathbb{E}(\tau^q)=O_q(1)$. Usando a desigualdade triangular em $L^K(\mathbb{Z}_N^{k-1})$, vemos que basta provar apenas que

$$\mathbb{E}\left(\tau(h\cdot(\widetilde{w}-\widetilde{w}'))^K\,\middle|\, h\in\mathbb{Z}_N^{k-1}\right)=O_K(1),$$

para todos $\widetilde{w},\widetilde{w}'\in\{0,1\}^{k-1}$ distintos entre si. Mas, sendo a transformação $h\mapsto h\cdot(\widetilde{w}-\widetilde{w}')$ um recobrimento, o lado esquerdo acima é $\mathbb{E}(\tau^K)=O_K(1)$. □

Voltemos agora a prova da proposição. Lembre que o lema 2.4.4 diz que as funções básicas anti-uniformes tomam valores no intervalo $I=[-2^{2^{k-1}},2^{2^{k-1}}]$. Pelo teorema de aproximação de Weierstrass, dado $\varepsilon>0$, podemos encontrar um polinômio P aproximando a função contínua Φ de modo que

$$\|\Phi(DF_1,\ldots,DF_K)-P(DF_1,\ldots,DF_K)\|_{L^\infty}\leq\varepsilon.$$

Como ν é uma medida (i.e., $\mathbb{E}(\nu)=1+o(1)$), temos

$$|\langle\nu-1,\Phi(DF_1,\ldots,DF_K)-P(DF_1,\ldots,DF_K)\rangle|\leq(2+o(1))\varepsilon.$$

Por outro lado, combinando os lemas 2.4.3, 2.4.5, obtemos que

$$|\langle\nu-1,P(DF_1,\ldots,DF_K)\rangle|=o_{K,\varepsilon}(1)$$

porque P depende apenas de K e ε. Fazendo N grande (dependendo de K,ε), vemos que

$$|\langle\nu-1,\Phi(DF_1,\ldots,DF_K)\rangle|\leq 4\varepsilon.$$

Isto finaliza a prova da proposição 2.4.1. □

Observação 2.4.5. *A única vez em todo livro em que aplicamos a condição de correlações foi no final da prova do lema 2.4.5.*

Resumo da subseção "Anti-Uniformidade":

Nesta subseção introduzimos o conceito de anti-uniformidade, o qual serve para identificar funções não uniformes (ou melhor, obstruções para a uniformidade). Com efeito, vimos que toda função F gera naturalmente uma função DF anti-uniforme básica tal que a correlação $\langle F, DF\rangle$ será grande sempre que F não for uniforme; além disso, vimos um resultado mostrando que a medida pseudo-aleatória se distribui bem com relação a *álgebra* gerada pelas funções anti-uniformes básicas.

A seguir, estudaremos as sigmas-álgebras geradas pelos conjuntos de nível de funções anti-uniforme, as quais são as peças básicas da sigma-álgebra com respeito a qual tomaremos esperanças de modo a obter funções boas (= uniformes).

2.4.3 Sigma-Álgebras geradas por funções anti-uniformes básicas

Veremos agora como funções anti-uniformes básicas geram naturalmente sigma-álgebras onde elas são bem comportadas (i.e. permitem o uso do teorema de Szemerédi na sua forma original).

Proposição 2.4.2. *Se ν é k-pseudoaleatória e $DF_1, \ldots, DF_K$ são funções anti-uniformes básicas. Para todo $\epsilon < 1$ e $\sigma < 1/2$ existe uma sigma-álgebra $\mathcal{B}$ tal que se N é um primo grande então:*

- $\|DF_j - \mathbb{E}(DF_j|\mathcal{B})\|_{L^\infty} \leq \epsilon \ \forall j$.
- *Existe um $\Omega \subset \mathcal{B}$ (conjunto excepcional) tal que $\mathbb{E}((\nu+1)1_\Omega) = O_{K,\epsilon}(\sigma^{1/2})$.*
- $\|(1-1_\Omega)\mathbb{E}(\nu-1|\mathcal{B})\|_{L^\infty} = O_{K,\epsilon}(\sigma^{1/2})$.

Demonstração. O ponto de partida da prova desta proposição é o seguinte lema garantindo que cada função gera uma sigma-álgebra:

Lema 2.4.6. *Seja ν uma medida k-pseudo-aleatória, $0 < \epsilon < 1$ e $0 < \sigma < 1/2$ parâmetros, e $G \in L^\infty(\mathbb{Z}_N)$ uma função tomando valores no intervalo $I := [-2^{2^{k-1}}, 2^{2^{k-1}}]$. Então, existe $\mathcal{B}_{\epsilon,\sigma}(G)$ uma sigma-álgebra tal que*

- *(G pertence a sua própria σ-álgebra) Para toda $\mathcal{B}$ σ-álgebra,*

$$\|G - \mathbb{E}(G|\mathcal{B} \vee \mathcal{B}_{\epsilon,\sigma}(G))\|_{L^\infty(\mathbb{Z}_N)} \leq \epsilon.$$

- *(Complexidade limitada) $\mathcal{B}_{\epsilon,\sigma}(G)$ tem $O(1/\epsilon)$ átomos.*

- *(Boa aproximação por funções contínuas de G) Se A é um átomo de $\mathcal{B}_{\epsilon,\sigma}(G)$, então existe $\Psi_A : I \to [0,1]$ tal que*

$$\|(1_A - \Psi_A(G))(\nu + 1)\|_{L^1(\mathbb{Z}_N)} = O(\sigma).$$

Mais ainda, Ψ_A pertence a um compacto $E \subset C^0(I)$ que independe de G, ν, N e A.

Prova do lema 2.4.6. Juntando o fato de ν ser uma medida com o teorema de Fubini, temos

$$\begin{aligned}\int_0^1 \sum_{n\in\mathbb{Z}} \mathbb{E}(1_{G(x)\in[\epsilon(n-\sigma+\alpha),\epsilon(n+\sigma+\alpha)]}(\nu(x)+1)|x \in \mathbb{Z}_N)d\alpha \\ = 2\sigma\mathbb{E}(1+\nu(x)|x \in \mathbb{Z}_N) = O(\sigma).\end{aligned}$$

Portanto, podemos fixar α tal que

$$\sum_{n\in\mathbb{Z}} \mathbb{E}(1_{G(x)\in[\epsilon(n-\sigma+\alpha),\epsilon(n+\sigma+\alpha)]}(\nu(x)+1)|x \in \mathbb{Z}_N) = O(\sigma). \quad (2.4.1)$$

Definimos $\mathcal{B}_{\epsilon,\sigma}(G)$ como a σ-álgebra cujos átomos são $G^{-1}([\epsilon(n+\alpha), \epsilon(n+\alpha+1)])$ para $n \in \mathbb{Z}$. Isto está bem-definido porque os intervalos $[\epsilon(n+\alpha), \epsilon(n+\alpha+1)]$ particionam a reta.

Claramente se $\mathcal{B}$ é uma σ-álgebra, então a função G restrita a um átomo de $\mathcal{B} \vee \mathcal{B}_{\epsilon,\sigma}(G)$ toma valores num intervalo de diâmetro ϵ, o que nos dá o primeiro item do lema (G pertence a sua própria σ-álgebra). Agora, seja $A := G^{-1}([\epsilon(n+\alpha), \epsilon(n+\alpha+1)])$ um átomo de $\mathcal{B}_{\epsilon,\sigma}(G)$. Como G toma valores em I, temos que $n = O(1/\epsilon)$ (caso contrário, $A = \emptyset$). Isto prova o segundo item do lema (complexidade limitada). Finalmente, seja $\psi_\sigma : \mathbb{R} \to [0,1]$ uma função corte fixada tal que $\psi_\sigma = 1$ em $[\sigma, 1-\sigma]$ e $\psi_\sigma = 0$ em $[-\sigma, 1+\sigma]$, e defina $\Psi_A(x) := \psi_\sigma(\frac{x}{\epsilon} - n - \alpha)$. Obviamente, Ψ_A varia num compacto $E_{\epsilon,\sigma}$ de $C^0(I)$ (pois n e α são limitados) e a igualdade (2.4.1) implica o terceiro item do lema (boa aproximação por funções contínuas de G). □

Agora voltamos a prova da proposição 2.4.2. Tomamos o seguinte $B := B_{\epsilon,\sigma}(DF_1) \vee \cdots \vee B_{\epsilon,\sigma}(DF_K)$, onde $B_{\epsilon,\sigma}(DF_j)$ é a sigma-álgebra do lema 2.4.6. Claramente o primeiro item da proposição 2.4.2 segue do primeiro item do lema 2.4.6. Por outro lado, como cada $B_{\epsilon,\sigma}(DF_j)$ tem $O(1/\epsilon)$ átomos, B é gerada por $O_{K,\epsilon}(1)$ átomos. Diremos que um átomo A de B é *pequeno* se $\mathbb{E}((\nu+1)1_A) \leq \sigma^{1/2}$. Denote por Ω a união de todos os átomos pequenos. Então $\Omega \in B$ e vale o segundo item da proposição 2.4.2. Para provar o último item, basta provar que

$$\frac{\mathbb{E}((\nu-1)1_A)}{\mathbb{E}(1_A)} = \mathbb{E}(\nu - 1|A) = o_{K,\epsilon,\sigma}(1) + O_{K,\epsilon}(\sigma^{1/2})$$

para todo átomo A não pequeno. Da definição de pequenez, temos que

$$\mathbb{E}((\nu-1)1_A) + 2\mathbb{E}(1_A) = \mathbb{E}((\nu+1)1_A) \geq \sigma^{1/2}$$

para A não pequeno. Logo, como σ é pequeno e N é grande, é suficiente verificar que

$$\mathbb{E}((\nu-1)1_A) = o_{K,\epsilon,\sigma}(1) + O_{K,\epsilon}(\sigma^{1/2}).$$

Por outro lado, sendo A a interseção de K átomos $A_j \in B_{\epsilon,\sigma}(DF_j)$, $j = 1, \ldots, K$, o lema 2.4.6 e a desigualdade de Hölder mostram que existe $\Psi_A : I^K \to [0,1]$ tal que

$$\|(\nu+1)(1_A - \Psi_A(DF_1, \ldots, DF_K))\|_{L^1(\mathbb{Z}_N)} = O_K(\sigma),$$

donde

$$\|(\nu-1)(1_A - \Psi_A(DF_1, \ldots, DF_K))\|_{L^1(\mathbb{Z}_N)} = O_K(\sigma).$$

Além disso, Ψ_A está num compacto $E_{\epsilon,K,\sigma}$ de $C^0(I^K)$. Isto e a proposição 2.4.1 (de distribuição uniforme com respeito a funções anti-uniformes básicas) implicam

$$\mathbb{E}((\nu-1)\Psi_A(DF_1, \ldots, DF_K)) = o_{K,\epsilon,\sigma}(1) = O_{K,\epsilon}(\sigma^{1/2}),$$

pois N é grande dependendo de K, ϵ, σ. Esta estimativa e a desigualdade triangular concluem a prova. □

Resumo da subseção "Sigma-Álgebras geradas por funções anti-uniformes básicas":

Nesta subseção associamos a cada função anti-uniforme básica DF uma *sigma-álgebra* $\mathcal{B}$ com respeito a qual a esperença $\mathbb{E}(DF|\mathcal{B})$ de DF aproxima DF (ou seja, DF é quase constante nos átomos de $\mathcal{B}$) e a medida pseudo-aleatória ν tem valores próximos a 1 em média (com relação a $\mathcal{B}$).

No estágio seguinte, usaremos esta maquinária de funções anti-uniformes básicas e suas sigma-álgebras para formalizar o processo de decomposição em partes boas (uniformes) e ruins (anti-uniformes) através de uma indução. Um ponto fundamental será garantir que este procedimento *pára* com um número finito de iterações. Isto seguirá do *argumento de incremento de energia*.

2.4.4 O argumento de incremento na energia

Usando as sigma-álgebras de funções anti-uniformes básicas, podemos obter a decomposição desejada em partes uniformes e partes anti-uniformes:

Teorema 2.4.2 (Koopman-von Neumann generalizado)**.** *Seja ν k-pseudoaleatória e $f \in L^1$ tal que $0 \leq f \leq \nu$, $\epsilon << 1$ e N é um primo grande. Então existe uma sigma-álgebra $\mathcal{B}$ e um conjunto excepcional $\Omega \in \mathcal{B}$ tal que:*

- $\mathbb{E}(\nu \cdot 1_\Omega) = o_\epsilon(1)$ *(o conjunto excepcional é pequeno).*
- $\|(1 - 1_\Omega)\mathbb{E}(\nu - 1|\mathcal{B})\|_{L^\infty} = o_\epsilon(1)$ *(boa distribuição da função fora do conjunto excepcional).*
- $\|(1 - 1_\Omega)(f - \mathbb{E}(f|\mathcal{B})\|_{U^{k-1}} \leq \epsilon^{1/2^k}$ *(uniformidade em $\mathcal{B}$)*

Demonstração. A estratégia básica é a mesma do teorema de estrutura ergódica de Furstenberg[7]: começamos com a sigma-álgebra $\mathcal{B} = \{\emptyset, \mathbb{Z}_N\}$ e olhamos para a função $f - \mathbb{E}(f|\mathcal{B})$. Se ela for uniforme (i.e., vale o terceiro item acima), acabamos. Caso contrário, usamos os resultados sobre anti-uniformidade para achar uma G_1

[7]Este teorema diz que podemos decompor qualquer sistema como uma extensão weak-mixing de uma torre de extensões compactas.

anti-uniforme com correlação não-trivial com f e adicionamos os conjuntos de nível de G_1 a sigma-álgebra $\mathcal{B}$. A propriedade de correlação não-trivial irá garantir que a norma L^2 de $\mathbb{E}(f|\mathcal{B})$ aumentará por uma quantidade não-trivial[8], enquanto que a pseudo-aleatoriedade de ν mostra que $\mathbb{E}(f|\mathcal{B})$ fica uniformemente limitado. Neste ponto, repetimos este procedimento até $f - \mathbb{E}(f|\mathcal{B})$ ficar suficientemente uniforme; note que o algoritmo irá parar em um número limitado de passo (da ordem de $2^{2^k}/\epsilon$) devido ao incremento de energia a cada passo.

Agora vamos escrever esta estratégia de modo um pouco mais organizado. Fixe ϵ e seja K_0 o menor inteiro maior que $1 + 2^{2^k}/\epsilon$. Precisaremos de um parâmetro $0 < \sigma \ll \epsilon$ e tomaremos N grande dependendo de ϵ e σ. Construiremos $\mathcal{B}$ e Ω através de uma sequência de funções anti-uniformes básicas $DF_1, \ldots, DF_K$ em $\mathbb{Z}_N$, conjuntos excepcionais $\Omega_0 \subset \cdots \subset \Omega_K \subset \mathbb{Z}_N$ e sigma-álgebras $\mathcal{B}_0 \subset \cdots \subset \mathcal{B}_K$ para algum $0 \leq K \leq K_0$ assim:

- Passo 0: Iniciamos com K=0, $\mathcal{B}_0 := \{\emptyset, \mathbb{Z}_N\}$ e $\Omega_0 := \emptyset$.
- Passo 1: Fazemos $F_{K+1} := (1 - 1_{\Omega_K})(f - \mathbb{E}(f|\mathcal{B}_K))$. Se
$$\|F_{K+1}\|_{U^{k-1}} \leq \epsilon^{1/2^k}$$
definimos $\Omega := \Omega_K$, $\mathcal{B} = \mathcal{B}_K$ e terminamos com sucesso o algoritmo.
- Passo 2: Caso
$$\|F_{K+1}\|_{U^{k-1}} > \epsilon^{1/2^k}$$
definimos $\mathcal{B}_{K+1} := \mathcal{B}_K \vee \mathcal{B}_{\epsilon,\sigma}(DF_{K+1})$.
- Passo 3: Procuramos por conjunto excepcional $\Omega_{K+1} \supset \Omega_K$ em $\mathcal{B}_{K+1}$ com
$$\mathbb{E}((\nu + 1)1_{\Omega_{K+1}}) = O_{K,\epsilon}(\sigma^{1/2}) \tag{2.4.2}$$
e
$$\|(1 - 1_{\Omega_{K+1}})\mathbb{E}(\nu - 1|\mathcal{B}_{K+1})\|_{L^\infty} = O_{K,\epsilon}(\sigma^{1/2}).$$
Se tal conjunto excepcional não puder ser achado, terminamos o algoritmo com erro. Caso contrário, vamos para o passo 4.

[8] A idéia de usar uma correlação não-trivial para aumentar a norma L^2 é precisamente o *argumento de incremento da energia*.

- Passo 4: Aumentamos K para $K+1$. Se $K > K_0$, terminamos o algoritmo com erro. Caso contrário, voltamos ao passo 1.

Assuma por enquanto que o algoritmo termina sem erro no passo 3 ou 4. Então é claro que após no máximo $K_0 + 1$ iterações, teremos construído uma σ-álgebra $\mathcal{B}$ e um conjunto excepcional Ω com as propriedades desejadas, exceto pelo fato de que os termos de erro são $O_{K,\epsilon}(\sigma^{1/2})$ ao invés de $o_\epsilon(1)$, para N grande dependendo de σ, K, ϵ. Entretanto, isto pode ser remediado fazendo σ tender a zero bem devagar.

Ou seja, reduzimos a prova deste teorema a mostrar que o algoritmo termina sem erro. A demonstração é por indução: como hipótese para indução em $0 \leq K_1 \leq K_0$, suponha que o algoritmo ou termina sem erro ou atinge o passo 2 da K_1-ésima iteração sem retornar um erro. Note que isto é obvio para $K_1 = 0$. Assumindo isto provado para algum $K_1 < K_0$, desejamos provar o mesmo para $K_1 + 1$. Observe que podemos supor que o algoritmo não terminou até o passo 2 do K_1-ésima iteração. Neste estágio, temos σ-álgebras $\mathcal{B}_0, \ldots, \mathcal{B}_{K_1+1}$, funções anti-uniformes básicas $DF_1, \ldots, DF_{K_1+1}$ e conjuntos excepcionais $\Omega_0, \ldots, \Omega_{K_1}$ já construídos. Afirmamos que

$$\|DF_j\|_{L^\infty} \leq 2^{2^{k-1}} + O_{j,\epsilon}(\sigma^{1/2}), \tag{2.4.3}$$

para todo $1 \leq j \leq K_1 + 1$. Isto segue do passo 3 das iterações anteriores (ou do passo 0 quando $j = 1$), os quais dizem que

$$\|(1 - 1_{\Omega_{j-1}})\mathbb{E}(\nu - 1|\mathcal{B}_{j-1})\|_{L^\infty} = O_{j,\epsilon}(\sigma^{1/2}),$$

donde

$$\mathbb{E}(\nu|\mathcal{B}_{j-1})(x) = 1 + O_{j-1,\epsilon}(\sigma^{1/2}),$$

para todo $x \notin \Omega_{j-1}$. Como $0 \leq f(x) \leq \nu(x)$, concluímos as estimativas pontuais

$$0 \leq (1 - 1_{\Omega_{j-1}}(x))\mathbb{E}(f|\mathcal{B}_{j-1})(x) \leq 1 + O_{j,\epsilon}(\sigma^{1/2}), \tag{2.4.4}$$

das quais seguem, por definição de F_j,

$$|F_j(x)| \leq (1 + O_{j,\epsilon}(\sigma^{1/2}))(\nu(x) + 1). \tag{2.4.5}$$

Em particular, uma simples aplicação do lema 2.4.4 prova a nossa afirmação acima.

Por outro lado, como $\mathcal{B}_{K_1+1}$ é a σ-álgebra gerada por

$$\mathcal{B}_{\epsilon,\alpha_1}(DF_1), \ldots, \mathcal{B}_{\epsilon,\alpha_{K_1+1}}(DF_{K_1+1}),$$

a proposição 2.4.2 permite encontrar Ω tal que

$$\mathbb{E}((\nu+1)1_\Omega) = O_{K_1,\epsilon}(\sigma^{1/2})$$

e

$$\|(1-1_\Omega)\mathbb{E}(\nu-1|\mathcal{B}_{K_1+1})\|_{L^\infty} = O_{K_1,\epsilon}(\sigma^{1/2}).$$

Definimos $\Omega_{K_1+1} := \Omega \cup \Omega_{K_1}$. Obviamente, Ω_{K_1+1} tem as propriedades necessárias para se executar o passo 3 sem erro, e portanto podemos ir até o passo 2 da K_1+1-ésima iteração (ou terminar sem erro), como queríamos provar.

Em outras palavras, o que provamos até agora foi que temos apenas duas possibilidades para o algoritmo: ou ele termina sem erro ou percorre todo o caminho até a K_0-ésima iteração. Para finalizar a demonstração, a propriedade-chave é que no caso do algoritmo atingir o passo 3 do K_0 iterado, então vale a estimativa de *incremento de energia*

$$\begin{aligned} &\|(1-1_{\Omega_j})\mathbb{E}(f|\mathcal{B}_j)\|_{L^2}^2 \\ &\geq \|(1-1_{\Omega_{j-1}})\mathbb{E}(f|\mathcal{B}_{j-1})\|_{L^2}^2 \\ &\quad + 2^{2^k-2}\epsilon - O_{j,\epsilon}(\sigma^{1/2}) - O(\epsilon^2), \end{aligned} \tag{2.4.6}$$

para todo $1 \leq j \leq K_0$ (se N é grande dependendo de K_0 e ϵ). Esta propriedade é suficiente para concluir a prova porque a desigualdade (2.4.4) nos dá

$$0 \leq \|(1-1_{\Omega_j})\mathbb{E}(f|\mathcal{B}_j)\|_{L^2}^2 \leq 1 + O_{j,\epsilon}(\sigma^{1/2}), \tag{2.4.7}$$

para todo $0 \leq j \leq K_0$. Como K_0 é o menor inteiro maior que $2^{2^k}/\epsilon + 1$, o príncipio da casa de pombos gera uma contradição para ϵ pequeno, σ pequeno e N grande dependendo de ϵ, σ.

Finalmente, resta apenas saber mostrar a estimativa de incremento na energia. A idéia é explorar o fato do algoritmo não parar no segundo passo da $(j-1)$-ésima iteração, o qual implica

$$\|F_j\|_{U^{k-1}} \geq \epsilon^{1/2^k}$$

Isto combinado com a definição de F_j e o lema 2.4.4 diz que

$$|\langle(1-1_{\Omega_{j-1}})(f-\mathbb{E}(f|\mathcal{B}_{j-1})), DF_j\rangle| = \|F_j\|_{U^{k-1}}^{2^{k-1}} \geq \epsilon^{1/2}.$$

Por outro lado, as estimativas pontuais (2.4.3), (2.4.5), (2.4.2) acima mostram que

$$\langle(1_{\Omega_j}-1_{\Omega_{j-1}})(f-\mathbb{E}(f|\mathcal{B}_{j-1}), DF_j)\rangle = O_{j,\epsilon}(\sigma^{1/2}),$$

enquanto que o lema 2.4.6 e a estimativa (2.4.5) falam que

$$\langle(1-1_{\Omega_j})(f-\mathbb{E}(f|\mathcal{B}_{j-1}), DF_j - \mathbb{E}(DF_j|\mathcal{B}_j))\rangle = O(\epsilon).$$

Em particular, pela desigualdade triangular, ganhamos a cota inferior:

$$|\langle(1-1_{\Omega_j})(f-\mathbb{E}(f|\mathcal{B}_{j-1})), \mathbb{E}(DF_j|\mathcal{B}_j)\rangle| \geq \epsilon^{1/2} - O_{j,\epsilon}(\sigma^{1/2}) - O(\epsilon).$$

Como as funções $(1-1_{\Omega_j})$, $\mathbb{E}(DF_j|\mathcal{B}_j)$ e $\mathbb{E}(f|\mathcal{B}_{j-1})$ são todas $\mathcal{B}_j$-mensuráveis, podemos trocar f por $\mathbb{E}(f|\mathcal{B}_j)$, de modo que obtemos

$$|\langle(1-1_{\Omega_j})(\mathbb{E}(f|\mathcal{B}_j)-\mathbb{E}(f|\mathcal{B}_{j-1})), \mathbb{E}(DF_j|\mathcal{B}_j)\rangle| \geq \epsilon^{1/2} - O_{j,\epsilon}(\sigma^{1/2}) - O(\epsilon).$$

Usando Cauchy-Schwarz e a estimativa (2.4.3) concluímos:

$$\|(1-1_{\Omega_j})(\mathbb{E}(f|\mathcal{B}_j)-\mathbb{E}(f|\mathcal{B}_{j-1}))\|_{L^2} \geq 2^{-2^{k-1}+1}\epsilon^{1/2} - O_{j,\epsilon}(\sigma^{1/2}) - O(\epsilon). \tag{2.4.8}$$

Esta estimativa moralmente implica, pelo teorema de Pitágoras, a estimativa de incremento de energia, o único problema sendo a presença dos conjuntos excepcionais Ω_{j-1}, Ω_j, os quais precisam de um pouco de cuidado para serem tratados, já que não temos boas cotas L^2 para ν. Para resolver este pequeno contra-tempo, começamos por notar que (2.4.2) e (2.4.4) implicam

$$\|(1_{\Omega_j}-1_{\Omega_{j-1}})\mathbb{E}(f|\mathcal{B}_{j-1})\|_{L^2} = O_{j,\epsilon}(\sigma^{1/2}).$$

Logo, a desigualdade triangular e (2.4.7) mostram que a estimativa de incremento de energia (2.4.6) segue de

$$\begin{aligned}&\|(1-1_{\Omega_j})\mathbb{E}(f|\mathcal{B}_j)\|_{L^2}^2 \\ &\geq \|(1-1_{\Omega_{j-1}})\mathbb{E}(f|\mathcal{B}_{j-1})\|_{L^2}^2 + \epsilon^{1/2} - O_{j,\epsilon}(\sigma^{1/2}) - O(\epsilon).\end{aligned}$$

Entretanto, o lado esquerdo acima pode ser expandido pela lei dos cossenos como

$$\|(1-1_{\Omega_j})\mathbb{E}(f|\mathcal{B}_{j-1})\|_{L^2}^2 + \|(1-1_{\Omega_j})(\mathbb{E}(f|\mathcal{B}_j)-\mathbb{E}(f|\mathcal{B}_{j-1}))\|_{L^2}^2 + 2\langle(1-1_{\Omega_j})\mathbb{E}(f|\mathcal{B}_{j-1}),(1-1_{\Omega_j})(\mathbb{E}(f|\mathcal{B}_j)-\mathbb{E}(f|\mathcal{B}_{j-1}))\rangle.$$

Portanto, por (2.4.8), vemos que é suficiente provar a seguinte relação de quase-ortogonalidade:

$$\langle(1-1_{\Omega_j})\mathbb{E}(f|\mathcal{B}_{j-1}),(1-1_{\Omega_j})(\mathbb{E}(f|\mathcal{B}_j)-\mathbb{E}(f|\mathcal{B}_{j-1}))\rangle = O_{j,\epsilon}(\sigma^{1/2}).$$

Como $(1-1_{\Omega_j})^2 = (1-1_{\Omega_j})$, podemos reescrever a identidade acima como

$$\langle(1-1_{\Omega_j})\mathbb{E}(f|\mathcal{B}_{j-1}),\mathbb{E}(f|\mathcal{B}_j)-\mathbb{E}(f|\mathcal{B}_{j-1})\rangle = O_{j,\epsilon}(\sigma^{1/2}).$$

Agora observemos que sendo a função $(1-1_{\Omega_{j-1}})\mathbb{E}(f|\mathcal{B}_{j-1})$ mensurável com relação a $\mathcal{B}_{j-1}$, ela deve ser ortogonal a função $\mathbb{E}(f|\mathcal{B}_j)-\mathbb{E}(f|\mathcal{B}_{j-1})$ porque $\mathcal{B}_{j-1}$ é uma sub-sigma-álgebra de $\mathcal{B}_j$ por construção. Em particular, podemos mais uma vez reescrever a expressão acima como

$$\langle(1_{\Omega_j}-1_{\Omega_{j-1}})\mathbb{E}(f|\mathcal{B}_{j-1}),\mathbb{E}(f|\mathcal{B}_j)-\mathbb{E}(f|\mathcal{B}_{j-1})\rangle = O_{j,\epsilon}(\sigma^{1/2}).$$

Usando novamente o fato de $(1_{\Omega_j}-1_{\Omega_{j-1}})\mathbb{E}(f|\mathcal{B}_{j-1})$ ser uma função $\mathcal{B}_j$-mensurável (donde segue que ela deve ser ortogonal a $f-\mathbb{E}(f|\mathcal{B}_j)$), vemos que a identidade acima equivale a

$$\langle(1_{\Omega_j}-1_{\Omega_{j-1}})\mathbb{E}(f|\mathcal{B}_{j-1}),f-\mathbb{E}(f|\mathcal{B}_{j-1})\rangle = O_{j,\epsilon}(\sigma^{1/2}).$$

Porém esta igualdade certamente é verdadeira porque $0 \leq f \leq \nu$ e valem as estimativas (2.4.2), (2.4.4). Isto completa a prova da estimativa de incremento de energia (2.4.6) e, consequentemente, a demonstração do teorema 2.4.2. □

Resumo da subseção "O argumento de incremento de energia":

Nesta subseção usamos toda a maquinária de sigma-álgebras associadas a funções anti-uniformes para exibir um *algoritmo* de construção

de conjuntos excepcionais *pequenos* e uma sigma-álgebra (para uma densidade f majorada por uma medida pseudo-aleatória) tais que que a função f possui um comportamento *uniforme* fora do conjunto excepcional. Em particular, isto nos mostra como decompor a função f em parte uniforme e parte não-uniforme. Este foi o conteúdo do teorema de Koopman-von Neumann generalizado 2.4.2. Mais ainda, o algoritmo levando a prova do teorema de Koopman-von Neumann generalizado era finito (i.e., ele parava em tempo finito) graças a um argumento de incremento de energia a cada passo (de fato, como a energia sempre crescia a cada passo e ela devia permanecer limitada durante todo processo, isto levava rapidamente a conclusão desejada).

O último passo será aplicar a decomposição fornecida pelo teorema 2.4.2 para finalizar a demonstração do teorema de Green-Tao-Szemerédi.

2.4.5 Fim da prova do teorema de Green-Tao-Szemerédi

Uma vez que já formalizamos (e quantificamos) toda a maquinária de uniformidade, anti-uniformidade e decomposição em partes uniforme e não-uniforme, a tarefa de imitar o esquema proposto na seção 2.3 para a prova do teorema de Roth visando demonstrar o teorema 2.2.1 de Green-Tao-Szemerédi é simples:

Sejam f, ν e δ como enunciado do teorema 2.2.1. Tome $\epsilon << \delta$ e considere $\mathcal{B}$ a sigma-álgebra do teorema 2.4.2 de Koopman-von Neumann generalizado. Defina as funções:

- $f_U = (1 - 1_\Omega)(f - \mathbb{E}(f|\mathcal{B}))$
- $f_{AU} = (1 - 1_\Omega)\mathbb{E}(f|\mathcal{B})$.

Lembre que, por hipotése, $0 \leq f \leq \nu$ (pontualmente) e $\mathbb{E}(f) \geq \delta$. Portanto, o teorema 2.4.2 garante que

$$\mathbb{E}(f_{AU}) = \mathbb{E}((1 - 1_\Omega)f) \geq \mathbb{E}(f) - \mathbb{E}(\nu 1_\Omega) \geq \delta - o_\epsilon(1).$$

Além disso, temos que $f_{AU} \leq 1 + o_\epsilon(1)$, de modo que podemos usar

o teorema de Szemerédi[9]. Em particular:

$$\mathbb{E}(f_{AU}(n)\dots f_{AU}(n+(k-1)r)|n,r\in\mathbb{Z}_N)\geq c(k,\delta)-o_\epsilon(1).$$

Por outro lado, sabemos que $\|f_U\|_{U^{k-1}}\leq\epsilon^{1/2^k}$. Logo, pelo teorema 2.4.1 de von Neumann generalizado, temos que:

$$\mathbb{E}(f_{*_1}(n)\dots f_{*_k}(n+(k-1)r)|n,r\in\mathbb{Z}_N)=O(\epsilon^{1/2^k})$$

onde $*_j=U$ ou AU, e $*_j=U$ para pelo menos um j.

Assim obtemos que:

$$\mathbb{E}(f(n)\dots f(n+(k-1)r)|n,r\in\mathbb{Z}_N)\geq c(k,\delta)-O(\epsilon^{1/2^k})-o_\epsilon(1).$$

Como ϵ é arbitrário, o teorema de Green-Tao-Szemerédi está provado!

Observação 2.4.6. *O leitor mais atento percebeu a analogia evidente entre as estimativas acima e a estimativas da proposição 2.3.1. De fato, como não podia deixar de ser, em ambos os argumentos, nós separamos o termo "bom" (no caso do Roth era $\Lambda_3(g,g,g)$ e no caso do Green-Tao era $\mathbb{E}(f_{AU}(n)\dots f_{AU}(n+(k-1)r)|n,r\in\mathbb{Z}_N)$) o qual é relativamente grande (no caso do Roth era δ^3 e no caso de Green-Tao é $c(k,\delta)-o_\epsilon(1)$) e ficamos por estimar os termos "ruins" (no caso do Roth eram $\Lambda_3(.,.,.)$ onde alguma das entradas tinha a função b e no caso de Green-Tao era $\mathbb{E}(f_{*_1}(n)\dots f_{*_k}(n+(k-1)r)|n,r\in\mathbb{Z}_N)$ onde alguma das entradas tinha a função f_U). Para cumprir esta tarefa, usamos Hölder no caso do Roth e o teorema de von Neumann generalizado no caso de Green-Tao para reduzir o problema ao "fato" (não-trivial) de que b no caso de Roth e f_U no caso de Green-Tao podiam ser escolhidas* uniformes. *Logicamente, este fato é obtido do enunciado do teorema de Koopman-von Neumann generalizado, o qual usa em sua prova o argumento de incremento de energia, conforme havíamos dito bem no ínicio.*

[9]Estamos fazendo uma pequena trapaça "inócula" aqui: como f_{AU} não é limitada por 1 exatamente e $\mathbb{E}(f_{AU})$ não é minorado por δ exatamente, o teorema de Szemerédi não pode ser usado diretamente. Porém, isso é facilmente contornado se aplicarmos Szemerédi a uma função é igual a f módulo um termo da forma $o_\epsilon(1)$, o qual pode ser trivialmente controlado nesse caso.

Capítulo 3

Construção da Medida Pseudo-Aleatória

3.1 A Medida Pseudo-Aleatória

Pretendemos provar neste capítulo o teorema 2.2.2, que recapitulamos a seguir (sob o nome de teorema 3.1.1):

Seja $\widetilde{\Lambda}$ a função de von Mangoldt modificada, dada por

$$\widetilde{\Lambda}(n) = \begin{cases} \dfrac{\phi(W)}{W} \log(Wn+1) \text{ se } Wn+1 \text{ é primo} \\ 0 \text{ caso contrário} \end{cases}$$

(onde $W = W(n) = \prod_{\substack{p \leq w(n) \\ p \text{ primo}}} p$; aqui $w(n)$ é uma função que tende lentamente a $+\infty$, mas observaremos no fim da prova que podemos tomar $w(n)$ uma constante grande).

Teorema 3.1.1. *Se $\varepsilon_k = 2^{-k}/(k+4)!$ e N é um primo grande, então existe uma medida k-pseudo-aleatória ν tal que $\nu(n) \geq 2^{-k-5}\, k^{-1}\, \widetilde{\Lambda}(n)$ para $\varepsilon_k N \leq n \leq 2\varepsilon_k N$.*

Vamos a seguir construir a medida ν. Sua construção e a prova de que realmente é uma medida pseudo-aleatória estão fortemente inspirados em resultados de Goldston e Yıldırım, principalmente [6].

A função $\widetilde{\Lambda}$ é uma modificação da clássica função de von Mangoldt dada por $\Lambda(n) = \begin{cases} \log p \text{ se } n = p^k, p \text{ primo}, k \geq 1 \\ 0 \text{ caso contrário} \end{cases}$. Essa modificação é feita para superar a falta de aleatoriedade do conjunto dos primos provocada por razões locais, i.e., pelo seu comportamento módulo primos pequenos.

Observamos agora que, se $n = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$, $p_1 < p_2 < \cdots < p_k$ primos, temos

$$\sum_{d|n} \Lambda(d) = \sum_{j=1}^{k} \sum_{r=1}^{\alpha_j} \Lambda(p_j^r) = \sum_{j=1}^{k} \sum_{r=1}^{\alpha_j} \log\, p_j = \sum_{j=1}^{k} \alpha_j \log p_j = \log n,$$

e logo, pela fórmula da inversão de Möbius (ver apêndice), $\Lambda(n) = \sum_{d|n} \mu(d) \log(n/d) = \sum_{d|n} \mu(d) \log_+(n/d)$, onde μ é a função de Möbius e $\log_+(x) = \max\{\log x, 0\}$, para $x > 0$. Isto motiva a seguinte definição, de Goldston e Yıldırım:

Definição 3.1.1 (Soma truncada sobre divisores de Goldston e Yıldırım)**.** *Seja $R > 0$ um parâmetro (nas aplicações será uma potência pequena de N).*

Definimos $\Lambda_R(n) = \sum_{\substack{d|n \\ d \leq R}} \mu(d) \log(R/d) = \sum_{d|n} \mu(d) \log_+(R/d)$.

Podemos agora definir a medida ν:

Definição 3.1.2. *Seja $R = N^{k^{-1}2^{-k-4}}$, e seja $\varepsilon_k = 2^{-k}/(k+4)!$. Definimos a função $\nu\colon \mathbb{Z}_N \to \mathbb{R}^+$ por*

$$\nu(n) = \begin{cases} \dfrac{\phi(W)}{W} \dfrac{\Lambda_R(Wn+1)^2}{\log R} & \text{se } \varepsilon_k N \leq n \leq 2\varepsilon_k N \\ 1, \text{ caso contrário} \end{cases}$$

Provar que ν é de fato uma medida pseudo-aleatória (ou mesmo uma medida) dará um certo trabalho. Entretanto é bastante simples mostrar que ν majora $\widetilde{\Lambda}$ como queremos:

Lema 3.1.1. $\nu(n) \geq k^{-1}2^{-k-5}\, \widetilde{\Lambda}(n)$ *para* $\varepsilon_k N \leq n \leq 2\varepsilon_k N$.

Demonstração. Isso é trivial se $Wn+1$ não é primo, pois nesse caso $\widetilde{\Lambda}(n) = 0$. Por outro lado, se $Wn+1$ é primo, os divisores de $Wn+1$ são apenas 1 e $Wn+1$ e, como W é grande, $Wn+1 > W\varepsilon_k N > R$ (assumindo, como sempre fazemos implicitamente, que N é suficientemente grande), e logo $\Lambda_R(Wn+1) = \sum_{\substack{d|Wn+1 \\ d\leq R}} \mu(d)\log(R/d) = \mu(1)\log R = \log R$, donde $\nu(n) = \dfrac{\phi(W)}{W}\log R$. Como $\widetilde{\Lambda}(n) = \dfrac{\phi(W)}{W}\log(Wn+1)$, e $\nu(n) = \dfrac{\phi(W)}{W}\log R = \dfrac{\phi(W)}{W}\cdot k^{-1}\cdot 2^{-k-4}\log N$, nossa afirmação equivale a $\log N \geq \dfrac{\log(Wn+1)}{2}$ para $\varepsilon_k N \leq n \leq 2\varepsilon_k N$, mas isso segue de $\log N \geq \dfrac{\log(WN+1)}{2}$, o que certamente é verdade se o crescimento de W é suficientemente lento ($W \leq N-1$ basta). □

Precisaremos de dois resultados seguintes, que provaremos posteriormente, os quais são essencialmente devidos a Goldston e Yıldırım, e que serão usados para provar que ν é uma medida pseudo-aleatória.

Proposição 3.1.1. *Sejam m e t inteiros positivos. Para cada $1 \leq i \leq m$, sejam $\psi_i(x) = \sum_{j=1}^{t} L_{ij}x_j + b_i$ formas lineares com coeficientes inteiros L_{ij} tais que $|L_{ij}| \leq \sqrt{w(N)}/2$ para $1 \leq i \leq m$ e $1 \leq j \leq t$. Assumimos que as t-uplas $(L_{ij})_{1\leq j\leq t}$ nunca são identicamente nulas, nem há duas dessas t-uplas que são múltiplos racionais uma da outra. Seja $\theta_i := W\psi_i + 1$. Suponha que B é um produto $\prod_{i=1}^{t} I_i \subset \mathbb{R}^t$ de t intervalos I_i, cada um tendo comprimento pelo menos R^{10m}. Então (desde que o crescimento da função $w(N)$ seja suficientemente lento)*

$$\mathbb{E}(\Lambda_R(\theta_1(x))^2 \dots \Lambda_R(\theta_m(x))^2 \mid x \in B) = (1 + o_{m,t}(1))\left(\frac{W\log R}{\phi(W)}\right)^m.$$

Proposição 3.1.2. *Seja $m \geq 1$ um inteiro, e seja B um intervalo de comprimento pelo menos R^{10m}. Suponha que $h_1, \dots, h_m$ sejam inteiros distintos tais que $|h_i| \leq m^2$ para $1 \leq i \leq m$ e seja $\Delta :=$*

$\prod_{1\le i<j\le m}|h_i-h_j|$. *Então (para N suficientemente grande dependendo de m, e supondo que o crescimento de $w(N)$ é suficientemente lento)*

$$\mathbb{E}(\Lambda_R(W(x_1+h_1)+1)^2\dots\Lambda_R(W(x_m+h_m)+1)^2 \mid x\in B)\le$$
$$\le (1+o_m(1))\left(\frac{W\log R}{\phi(W)}\right)^m \prod_{\substack{p\mid\Delta\\ p\ primo}}(1+O_m(p^{-1/2})).$$

Em geral, no que segue, usaremos a letra p sempre para denotar primos, como aqui.

Vamos inicialmente mostrar como concluir a prova do Teorema 3.1.1 a partir das Proposições 3.1.1 e 3.1.2 para depois demonstrar as proposições. Começaremos mostrando que ν é de fato uma medida.

Lema 3.1.2. *A medida ν construída na Definição 3.1.2 satisfaz a estimativa $\mathbb{E}(\nu)=1+o(1)$.*

Demonstração. Aplicamos a Proposição 3.1.1 com $m=t=1$, $\psi_1(x_1)=x_1$ e $B=[\varepsilon_k N, 2\varepsilon_k N]$. Comparando com a Definição 3.1.2 temos $\mathbb{E}(\nu(x)\mid x\in[\varepsilon_k N,2\varepsilon_k N])=1+o(1)$, pois a Proposição 3.1.1 fornece $\mathbb{E}(\Lambda_R(Wx+1)^2\mid x\in B)=(1+o(1))\cdot\dfrac{W\log R}{\phi(W)}\cdot$ Por outro lado, obviamente temos $\mathbb{E}(\nu(x)\mid x\in\mathbb{Z}_N\backslash[\varepsilon_k N,2\varepsilon_k N])=1$, pela mesma Definição 3.1.2.

Combinando os dois resultados concluímos a prova do lema. □

Proposição 3.1.3. *A função ν satisfaz a $(k\cdot 2^{k-1}, 3k-4, k)$ – condição de formas lineares.*

Demonstração. Sejam $\psi_i(x)=\sum_{j=1}^{t}L_{ij}+b_i$ formas lineares como na Definição 2.2.1, isto é, temos $m\le k\cdot 2^{k-1}$, $t\le 3k-4$, e os L_{ij} são números racionais com numerador e denominador de valor absoluto no máximo k de modo que nenhuma das t-uplas $(L_{ij})_{1\le j\le t}$ é nula ou múltiplo racional de alguma outra. Queremos mostrar que

$$\mathbb{E}(\nu(\psi_1(x))\dots\nu(\psi_m(x))\mid x\in\mathbb{Z}_N^m)=1+o(1). \qquad (*)$$

Podemos supor que os L_{ij} são inteiros, compondo $\psi_1, \psi_2, \ldots, \psi_m$ com a multiplicação por $k!$ dada por $M\colon \mathbb{Z}_M^m \to \mathbb{Z}_N^m$, $M(x) = k!x (\mathrm{mod}\ N)$, $\forall\, x \in \mathbb{Z}_N^m$. Como $\mathrm{mdc}(k!, N) = 1$, para N grande, M é uma bijeção, e o valor esperado não muda. Com isso, a cota superior para os $|L_{ij}|$ muda para $k \cdot k! < (k+1)!$. Como $w(N)$ tende a infinito quando N cresce, podemos supor que $(k+1)! < \sqrt{w(N)}/2$, tomando N grande. Precisamos deste fato para poder aplicar a Proposição 3.1.1.

Como a definição de ν tem duas partes, não é possível aplicar diretamente a Proposição 3.1.1. Vamos então dividir $\mathbb{Z}_N^m$ em Q^t blocos de lados quase iguais, onde $Q = Q(N)$ é uma função de crescimento lento em N a ser escolhida posteriormente. Sejam então os blocos

$$B_{u_1\, u_2 \ldots u_t} = \{x \in \mathbb{Z}_N^m\,;\, x_j \in [Lu_jN/Q, L(u_j+1)N/Q])\,,\ 1 \le j \le t\}$$

onde $u_1, u_2, \ldots, u_t \in \mathbb{Z}_Q$, e identificamos $\mathbb{Z}_Q$ com $\{0, 1, 2, \ldots, Q-1\}$.

Note que módulo um erro multiplicativo de $1 + o(1)$, podemos escrever o lado esquerdo de (*) como

$$\mathbb{E}(\mathbb{E}(\nu(\psi_1(x)) \ldots \nu(\psi_m(x)) \mid x \in B_{u_1\, u_2 \ldots u_t}) \mid u_1, \ldots, u_t \in \mathbb{Z}_Q).$$

Chamamos uma t-upla $(u_1, u_2, \ldots, u_t) \in \mathbb{Z}_Q^t$ boa se para $1 \le i \le m$, cada conjunto $\psi_i(B_{u_1 u_2 \ldots u_t})$ está completamente contido no intervalo $[\varepsilon_k N, 2\varepsilon_k N]$ ou é completamente disjunto deste intervalo. Tomando a Proposição 3.1.1 e a Definição 3.1.2, observamos que $\mathbb{E}(\nu(\psi_1(x)) \ldots \nu(\psi_m(x)) \mid x \in B_{u_1 u_2 \ldots u_k}) = 1 + o_{m,t}(1)$, sempre que $(u_1, u_2, \ldots, u_t)$ for boa, pois podemos substituir cada fator $\nu(\psi_i(x))$ por $\dfrac{\phi(W)}{W \log R}\, \Lambda_R^2(\theta_i(x))$ ou por 1, e, se o crescimento de Q é suficientemente lento, $N/Q > R^{10m}$, pela definição de R e pela limitação de m.

Se $(u_1, u_2, \ldots, u_t)$ não é boa, podemos majorar ν por $1 + \dfrac{\phi(W)}{W \log R}\, \Lambda_R^2(\theta_i(x))$, multiplicar, expandir e aplicar a Proposição 3.1.1 a cada termo para obter uma cota do tipo

$$\mathbb{E}(\nu(\psi_1(x)) \ldots \nu(\psi_m(x)) \mid x \in B_{u_1 u_2 \ldots u_k}) = \mathcal{O}_{m,t}(1).$$

Veremos agora que a proporção de t-uplas $(u_1, u_2, \ldots, u_t) \in \mathbb{Z}_Q^t$ que não são boas é no máximo $\mathcal{O}_{m,t}(1/Q)$, e logo o lado direito de (*)

é $1+o_{m,t}(1)+\mathcal{O}_{m,t}(1/Q) = 1+o_{m,t}(1)$, o que conclui a demonstração. Para isso, suponha que $(u_1, u_2, \dots, u_t)$ não é boa. Então existem $i \leq m$ e $x, x' \in B_{u_1 u_2 \dots u_t}$ tais que $\psi_i(x)$ pertence ao intervalo $[\varepsilon_k N, 2\varepsilon_k N]$ mas $\psi_i(x')$ não. Pela definição de $B_{u_1 u_2 \dots u_t}$ (e pela limitação dos L_{ij}), temos $\psi_i(x)$, $\psi_i(x') = \sum_{j=1}^{t} L_{ij}\lfloor Nu_j/Q \rfloor + b_i + \mathcal{O}_{m,t}(N/Q)$. Portanto, para algum $a \in \{1, 2\}$, $a\,\varepsilon_k N = \sum_{j=1}^{t} L_{ij}\lfloor Nu_j/Q \rfloor + b_i + \mathcal{O}_{m,t}(N/Q)$. Dividindo por N/Q, obtemos $\sum_{j=1}^{t} L_{ij}u_j = a\,\varepsilon_k Q + b_i\, Q/N + \mathcal{O}_{m,t}(1)$ (mod Q). Como $(L_{ij})_{1\leq j\leq t}$ é não-nulo, o número de t-uplas $(u_1, u_2, \dots, u_t)$ que safisfazem esta equação é no máximo $\mathcal{O}_{m,t}(Q^{t-1})$. Fazendo a e i variarem, concluímos que a proporção de t-uplas que não são boas é no máximo $\mathcal{O}_{m,t}(1/Q)$, como queríamos. □

Veremos a seguir como usar a Proposição 3.1.2 para mostrar que ν satisfaz a condição de correlações. Para isso, vamos incialmente estimar o fator $\prod_{\substack{p|\Delta \\ p \text{ primo}}} (1 + \mathcal{O}_m(p^{-1/2}))$ que aparece na proposição:

Lema 3.1.3. *Seja $m \geq 1$ um parâmetro. Existe uma função peso $\tau = \tau_m \colon \mathbb{Z} \to \mathbb{R}^+$ tal que $\tau(n) \geq 1$, $\forall\, n \neq 0$ e, para cada $h_1, h_2, \dots, h_m \in [\varepsilon_k N, 2\varepsilon_k N]$ distintos, temos*

$$\prod_{\substack{p|\Delta \\ p\ primo}} (1 + \mathcal{O}_m(p^{-1/2})) \leq \sum_{1\leq i<j\leq m} \tau(h_i - h_j),$$

onde Δ foi definido na Proposição 3.1.2, de modo que $\mathbb{E}(\tau^q(n) \mid 0 < |n| \leq N) = \mathcal{O}_{m,q}(1)$, para $0 < q < \infty$.

Demonstração. Note que

$$\prod_{\substack{p|\Delta \\ p \text{ primo}}} (1 + \mathcal{O}_m(p^{-1/2})) \leq \prod_{1\leq i<j\leq m} \left(\prod_{\substack{p|h_i-h_j \\ p \text{ primo}}} (1 + p^{-1/2}) \right)^{\mathcal{O}_m(1)}.$$

Podemos então, usando a desigualdade entre as médias aritmética e geométrica (e absorvendo as constantes no fator e no expoente $\mathcal{O}_m(1)$), tomar $\tau_m(n) = \mathcal{O}_m(1) \prod_{\substack{p|n \\ p \text{ primo}}} (1+\mathcal{O}_m(p^{-1/2}))^{\mathcal{O}_m(1)}$ para cada $n \neq 0$ (o valor de τ em 0 é irrelevante para o lema, pois estamos tomando os h_j distintos).

Para concluir a prova do lema, basta mostrar que

$$\mathbb{E}\Big(\prod_{\substack{p|n \\ p \text{ primo}}} (1+p^{1/2})^{\mathcal{O}_n(1)} \mid 0 < |n| \leq N \Big) = \mathcal{O}_{m,q}(1), \text{ para } 0 < q < \infty.$$

Como $(1 + p^{-1/2})^{\mathcal{O}_m(q)} \leq 1 + p^{-1/4}$ para todos os primos p, com exceção de no máximo $\mathcal{O}_{m,q}(1)$ primos, temos

$$\begin{aligned} &\mathbb{E}\Big(\prod_{\substack{p|n \\ p \text{ primo}}} (1 + p^{-1/2})^{\mathcal{O}_m(q)} \mid 0 < |n| \leq N \Big) \\ &= \mathcal{O}_{m,q}(1) \cdot \mathbb{E}\Big(\prod_{\substack{p|n \\ p \text{ primo}}} (1 + p^{-1/4}) \mid 0 < n \leq N \Big). \end{aligned}$$

Por outro lado, $\prod_{\substack{p|n \\ p \text{ primo}}} (1 + p^{-1/4}) \leq \sum_{d|n} d^{-1/4}$, e logo

$$\begin{aligned} &\mathbb{E}\Big(\prod_{\substack{p|n \\ p \text{ primo}}} (1 + p^{-1/2})^{\mathcal{O}_m(q)} \mid 0 < |n| \leq N \Big) \\ &\leq \mathcal{O}_{m,q}(1) \cdot \frac{1}{N} \sum_{1 \leq n \leq N} \sum_{d|n} d^{-1/4} \\ &\leq \mathcal{O}_{m,q}(1) \cdot \frac{1}{N} \sum_{d=1}^{N} \frac{N}{d} \cdot d^{-1/4} = \mathcal{O}_{m,q}(1), \end{aligned}$$

pois $\sum_{d=1}^{\infty} d^{-5/4} < \infty$. □

Podemos agora provar a condição de correlações.

Proposição 3.1.4. *A medida ν satisfaz a 2^{k-1}-condição de correlações.*

Demonstração. Queremos mostrar que, para $1 \leq m \leq 2^{k-1}$ e $h_1, \ldots, h_m \in \mathbb{Z}_N$ temos $\mathbb{E}(\nu(x+h_1)\ldots\nu(x+h_m) \mid x \in \mathbb{Z}_N) \leq \sum_{1\leq i<j\leq m} \tau(h_i - h_j)$, onde a função peso $\tau = \tau_m$ é limitada em L^q para todo q.

Fixemos $m, h_1, \ldots, h_m$. Vamos tomar a função peso construída no Lema 3.1.3 (identificando $\mathbb{Z}_N$ com os inteiros entre $-N/2$ e $N/2$) multiplicada por um fator constante $\mathcal{O}_m(1)$ conveniente e definir $\tau(0) = \exp(Cm \log N/\log\log N)$, para alguma constante absoluta grande C. Pelo lema anterior concluímos que $\mathbb{E}(\tau^q) = \mathcal{O}_{m,q}(1)$ para todo q, pois $\tau(0)$ só contribui com $o_{m,q}(1)$ para o valor de $\mathbb{E}(\tau^q)$. Trataremos inicialmente do caso em que dois dos h_i são iguais. Nesse caso, podemos usar a estimativa grosseira $||\nu||_{L^\infty} \leq \exp\left(\dfrac{2\log N}{\log\log N}\right)$, que segue da definição de ν (na verdade obtemos facilmente da definição que $|\nu(n)| \leq \log N \cdot d(n)^2$, onde $d(n)$ é o número de divisores de n; temos, por outro lado, $d(n) = \mathcal{O}\left(\exp\left(\dfrac{3\log N}{4\log\log N}\right)\right)$ – ver apêndice, donde segue nosssa estimativa), e a afirmação nesse caso segue da escolha de $\tau(0)$.

Suponhamos agora que os h_i são todos distintos. Seja

$$g(n) := \frac{\phi(W)}{W} \cdot \frac{\Lambda_R^2(Wn+1)}{\log R} \cdot 1_{[\varepsilon_k N, 2\varepsilon_k N]}(n).$$

Pela construção de ν, temos

$$\mathbb{E}(\nu(x+h_1)\ldots\nu(x+h_m) \mid x \in \mathbb{Z}_N)$$

$$\leq \mathbb{E}((1+g(x+h_1))\ldots(1+g(x+h_m)) \mid x \in \mathbb{Z}_N).$$

O lado direito acima pode ser reescrito como

$$\sum_{A\subset\{1,\ldots,m\}} \mathbb{E}\left(\prod_{i\in A} g(x+h_i) \mid x \in \mathbb{Z}_N\right).$$

Note que para $i, j \in A$ podemos supor que $|h_i - h_j| \leq \varepsilon_k N$ (pois, caso contrário, a esperança correspondente se anula). Pela Proposição 3.1.2 e pelo Lema 3.1.3 obtemos portanto

$$\mathbb{E}\Big(\prod_{i\in A} g(x+h_i) \mid x \in \mathbb{Z}_N \Big) \leq (1+\mathcal{O}_m(1)) \sum_{1\leq i<j\leq m} \tau_m(h_i - h_j),$$

e somando sobre todos os A, obtemos o resultado, após multiplicar τ_m por uma constante adequada. □

Os Lemas 3.1.1 e 3.1.2, e as Proposições 3.1.3 e 3.1.4 concluem a prova de que ν é uma medida pseudo-aleatória e do Teorema 3.1.1.

Vamos agora nos dedicar a provar as Proposições 3.1.1 e 3.1.2.

3.2 Condição de formas lineares para Λ_R

Vamos provar a Proposição 3.1.1. Lembramos que temos, para cada i com $1 \leq i \leq m$, uma forma linear $\psi_i(x) = \sum_{j=1}^{t} L_{ij}\, x_j + b_i$ em t variáveis $x_1, \ldots, x_t$. Os coeficientes L_{ij} satisfazem $|L_{ij}| \leq \sqrt{w(N)}/2$, onde $w(N)$ é uma função com crescimento lento em N. Supomos que nenhum t-upla $(L_{ij})_{1\leq j\leq t}$ é nula ou múltiplo de alguma outra. Definimos $\theta_i = W\psi_i + 1$. Seja $B = \prod_{j=1}^{t} I_j$ um produto de intervalos I_j, cada um com comprimento maior ou igual a R^{10m}. Queremos provar a estimativa

$$\mathbb{E}(\Lambda_R(\theta_1(x))^2 \ldots \Lambda_R(\theta_n(x))^2 \mid x \in B) = (1+o(1)) \Big(\frac{W \log R}{\phi(W)} \Big)^m.$$

A primeira etapa da prova será eliminar o papel do bloco B. Podemos usar a definição de Λ_R para expandir o lado esquerdo como

$$\mathbb{E}\Big(\prod_{i=1}^{m} \sum_{\substack{d_i, d_i' \leq R \\ d_i, d_i' \mid \theta_i(x)}} \mu(d_i)\mu(d_i') \log \frac{R}{d_i} \log \frac{R}{d_i'} \mid x \in B \Big),$$

o que pode ser reescrito como

$$\sum_{d_1,\dots,d_m,d'_1,\dots,d'_m\le R}\Big(\prod_{i=1}^{m}\mu(d_i)\mu(d'_i)\log\frac{R}{d_i}\log\frac{R}{d'_i}\Big)\times \\ \times\,\mathbb{E}\Big(\prod_{i=1}^{m}\mathbf{1}_{d_i,d'_i|\theta_i(x)}\mid x\in B\Big). \tag{3.1}$$

Devido à presença da função μ de Möbius podemos supor que os d_i, d'_i são livres de quadrados. Seja $D=\operatorname{mmc}(d_1,\dots,d_m,d'_1,\dots,d'_m)$ o mínimo múltiplo comum dos d_i e dos d'_i; temos $D\le R^{2m}$. Note que a expressão $\prod_{i=1}^{m}\mathbf{1}_{d_i,d'_i|\theta_i(x)}$ é periódica com período D em cada coordenada de x, e portanto pode ser definido com domínio $\mathbb{Z}_D^t$. Como B é um produto de intervalos de comprimento maior ou igual a R^{10m}, temos

$$\mathbb{E}\Big(\prod_{i=1}^{m}\mathbf{1}_{d_i,d'_i|\theta_i(x)}\mid x\in B\Big) \\ =\mathbb{E}\Big(\prod_{i=1}^{m}\mathbf{1}_{d_i,d'_i|\theta_i(x)}\mid x\in\mathbb{Z}_D^t\Big)+\mathcal{O}_{m,t}(R^{-8m}).$$

A contribuição dos termos de erro $\mathcal{O}_{m,t}(R^{-8m})$ para (3.1) pode ser majorada grosseiramente por $R^{2m}(\log R)^{2m}\cdot\mathcal{O}_{m,t}(R^{-8m})=\mathcal{O}_{m,t}(R^{-6m}(\log R)^{2m})$. Basta mostrar, portanto, que

$$\sum_{d_1,\dots,d_m,d'_1,\dots,d'_m\le R}\Big(\prod_{i=1}^{m}\mu(d_i)\mu(d'_i)\log\frac{R}{d_i}\log\frac{R}{d'_i}\Big)\times \\ \times\,\mathbb{E}\Big(\prod_{i=1}^{m}\mathbf{1}_{d_i,d'_i|\theta_i(x)}\mid x\in\mathbb{Z}_D^t\Big) \\ =(1+o(1))\Big(\frac{W\log R}{\phi(W)}\Big)^m. \tag{3.2}$$

Para provar isso, escreveremos o lado esquerdo como uma integral de linha de um produto de Euler, que por sua vez pode ser escrito em termos da função ζ de Riemann e outros fatores simples.

Começamos usando o teorema chinês dos restos (e o fato dos d_i, d_i' serem livres de quadrados) para reescrever

$$\mathbb{E}\Big(\prod_{i=1}^{m} \mathbf{1}_{d_i,d_i'|\theta_i(x)} \mid x \in \mathbb{Z}_D^t\Big)$$
$$= \prod_{\substack{p|D \\ p \text{ primo}}} \mathbb{E}\Big(\prod_{i,p|d_i d_i'} \mathbf{1}_{\theta_i(x)\equiv 0(\text{mod } p)} \mid x \in \mathbb{Z}_p^t\Big)$$

(o lado esquerdo é a probabilidade de que $\theta_i(x)$ seja múltiplo de d_i e de d_i', para todo $i \leq m$, o que equivale a θ_i ser múltiplo de p sempre que p for um primo que divide d_i ou d_i', pois os d_i, d_i' são livres de quadrados). Note que a restrição $p|D$ no lado direito pode ser removida, pois, se $p \nmid D$, p nunca dividirá $d_i d_i'$, donde o multiplicando nesse caso é 1. Assim, escrevendo $X_{d_1,\ldots,d_m}(p) := \{1 \leq i \leq m; p \mid d_i\}$ e

$$\omega_X(p) := \mathbb{E}\Big(\prod_{i\in X} \mathbf{1}_{\theta_i(x)\equiv 0(\text{mod } p)} \mid x \in \mathbb{Z}_p^t\Big),$$

para cada $X \subset \{1, 2, \ldots, m\}$, temos

$$\mathbb{E}\Big(\prod_{i=1}^{m} \mathbf{1}_{d_i,d_i'|\theta_i(x)} \mid x \in \mathbb{Z}_D^t\Big) = \prod_{p \text{ primo}} \omega_{X_{d_1,\ldots,d_m}(p)\cup X_{d_1',\ldots,d_m'}(p)}(p).$$

Podemos então escrever o lado esquerdo de (3.2) como

$$\sum_{d_1,\ldots,d_m,d_1',\ldots,d_m' \in \mathbb{Z}^+} \Big(\prod_{i=1}^{m} \mu(d_i)\mu(d_i) \log_+\big(\frac{E}{d_i'}\big) \log_+\big(\frac{E}{d_i'}\big)\Big) \times$$
$$\times \prod_{p \text{ primo}} \omega_{X_{d_1,\ldots,d_m}(p)\cup X_{d_1',\ldots,d_m'}(p)}(p).$$

A seguir, vamos expressar os logaritmos em termos de funções multiplicativas dos d_i, d_i' por meio de integrais de linha. Para isso, usaremos o seguinte resultado:

Lema 3.2.1. *Dado $\varepsilon > 0$, seja $\Gamma(t)$ a reta vertical parametrizada por $\Gamma(t) = \varepsilon + it$, $-\infty < t < +\infty$. Temos então, para cada $x > 0$ real,* $\dfrac{1}{2\pi i}\displaystyle\int_\Gamma \frac{x^z}{z^2}\,dz = \log_+(x)$.

Demonstração. Dado $M > \varepsilon$, seja $\Gamma^{(M)}$ a restrição de Γ ao intervalo $\big[-\sqrt{M^2-\varepsilon^2}, \sqrt{M^2-\varepsilon^2}\big]$. Se $0 < x < 1$ considere o caminho fechado $\widetilde{\Gamma}^{(M)}$ formado por Γ^M seguido do caminho

$$\widetilde{\gamma}^{(M)}\colon [-\cos^{-1}(\varepsilon/M), \cos^{-1}(\varepsilon/M)] \to \mathbb{C},\ \widetilde{\gamma}^{(M)}(t) = Me^{-it}.$$

Como a função x^z/z^2 não tem singularidades no interior de $\widetilde{\Gamma}^{(M)}$, temos $\displaystyle\int_{\widetilde{\Gamma}^{(M)}} \frac{x^z}{z^2}\,dz = 0$. Por outro lado, como $0 < x < 1$, $\big|\displaystyle\int_{\widetilde{\gamma}^{(M)}} \frac{x^z}{z^2}\,dz\big| = \mathcal{O}\big(\frac{1}{M}\big)$, donde, nesse caso,

$$\begin{aligned}\frac{1}{2\pi i}\int_\Gamma \frac{x^z}{z^2}dz &= \lim_{M\to\infty}\frac{1}{2\pi i}\int_{\Gamma^{(M)}}\frac{x^z}{z^2}dz\\ &= -\lim_{M\to\infty}\frac{1}{2\pi i}\int_{\widetilde{\gamma}^{(M)}}\frac{x^z}{z^2}dz = 0 = \log_+(x).\end{aligned}$$

Se $x \geq 1$, considere o caminho fechado $\widehat{\Gamma}^{(M)}$ formado por $\Gamma^{(M)}$ seguido do caminho $\hat{\gamma}^{(M)}\colon [\cos^{-1}(\varepsilon/M), 2\pi - \cos^{-1}(\varepsilon/M)] \to \mathbb{C}$, $\hat{\gamma}^{(M)}(t) = Me^{it}$. A única singularidade de x^z/z^2 no interior de $\widehat{\Gamma}^{(M)}$ é $z = 0$, e $x^z/z^2 = e^{z\log x}/z^2 = \dfrac{1 + z\log x + z^2\log^2 x/2 + \dots}{z^2} = z^{-2} + z^{-1}\log x + \dfrac{\log^2 x}{2} + \dots$ tem resíduo $\log x$ em $z = 0$, donde

$$\frac{1}{2\pi i}\int\limits_{\widehat{\Gamma}^{(M)}} x^z/z^2\,dz = \log x.$$

Como $x \geq 1$ e $\operatorname{Re} z \leq \varepsilon$ em $\hat{\gamma}^{(M)}$, $\left|\displaystyle\int\limits_{\hat{\gamma}^{(M)}} x^z/z^2\,dz\right| = \mathcal{O}\big(\frac{1}{M}\big)$, donde, nesse caso,

$$\begin{aligned}\frac{1}{2\pi i}\int_\Gamma \frac{x^z}{z^2}\,dz &= \lim_{M\to\infty}\frac{1}{2\pi i}\int_{\Gamma^{(M)}}\frac{x^z}{z^2}\,dz =\\ &= \log x - \lim_{M\to\infty}\int_{\hat{\gamma}^{(M)}}\frac{x^z}{z^2}\,dz = \log x = \log_+(x).\end{aligned}$$

□

A seguir, $\Gamma_1(t)$ denotará a reta vertical $\Gamma(t)$ correspondente a $\varepsilon = \dfrac{1}{\log R}$, isto é, $\Gamma_1(t) := \dfrac{1}{\log R} + it$, $-\infty < t < \infty$. Temos, pelo lema anterior,

$$\frac{1}{2\pi i}\int_{\Gamma_1} \frac{x^z}{z^2}\, dz = \log_+(x).$$

Note que R^z é limitado em Γ_1: $\left|R^{\Gamma_1(t)}\right| = R^{1/\log R} = e$, $\forall\, t \in \mathbb{R}$. Podemos, usando a identidade acima, reescrever o lado esquerdo de (3.2) como

$$(2\pi i)^{-2m} \int_{\Gamma_1} \dots \int_{\Gamma_1} F(z,z') \prod_{j=1}^{m} \frac{R^{z_j+z'_j}}{z_j^2 z_j'^2}\, dz_j\, dz'_j\,,$$

onde há $2m$ integrais de linha nas variáveis $z_1, \dots, z_m$, $z'_1, \dots, z'_m$ em Γ_1, $z := (z_1, \dots, z_m)$, $z' := (z'_1, \dots, z'_m)$ e

$$\begin{aligned} F(z,z') := \sum_{d_1,\dots,d_m,d'_1,\dots,dm' \in \mathbb{Z}^+} \left(\prod_{j=1}^{m} \frac{\mu(d_j)\mu(d'_j)}{d_j^{z_j} d_j'^{z'_j}}\right) \times \\ \times \prod_{p \text{ primo}} \omega_{X_{d_1,\dots,d_m}(p) \cup X_{d'_1,\dots,d'_m}(p)}(p). \end{aligned} \tag{3.3}$$

Observe que o somando em (3.3) é uma função multiplicativa dos d_j, d'_j, e portanto temos (pelo menos formalmente) uma representação em produto de Euler $F(z,z') = \prod\limits_{p \text{ primo}} E_p(z,z')$, onde

$$E_p(z,z') := \sum_{X,X' \subset \{1,\dots,m\}} \frac{(-1)^{|X|+|X'|}\, \omega_{X \cup X'}(p)}{p^{\sum\limits_{j\in X} z_j + \sum\limits_{j \in X'} z'_j}}\,.$$

Da definição de $\omega_X(p)$ temos $\omega_\phi(p) = 1$ e $\omega_X(p) \le 1$, donde $E_p(z,z') = 1 + \mathcal{O}_\sigma(1/p^\sigma)$ quando $\mathrm{Re}(z_j), \mathrm{Re}(z'_j) > \sigma$. Portanto o produto de Euler acima é absolutamente convergente (e vale a igualdade em (3.3)) no domínio $\{\mathrm{Re}(z_j), \mathrm{Re}(z'_j) \ge 1\}$, pelo menos.

Vamos agora explorar a hipótese sobre as partes lineares de $\psi_1, \dots, \psi_m$ serem não-nulas e não serem múltiplos racionais de nenhuma outra.

Lema 3.2.2 (Estimativa do fator local)**.** *Se $p \le w(N)$, então $\omega_X(p) = 0$ para todo conjunto não-vazio X. Em particular, $E_p = 1$ se $p \le w(N)$. Se $p > w(N)$, então $\omega_X(p) = p^{-1}$ quando $|X| = 1$ e $\omega_X(p) \le p^{-2}$ quando $|X| \ge 2$.*

Demonstração. A primeira afirmação é imediata, pois as funções $\theta_j \colon \mathbb{Z}_p^t \to \mathbb{Z}_p$ são identicamente 1 quando $p \le w(N)$. Para a segunda afirmação, observe que se $p > w(N)$ e $X = \{j\}$, cada elemento de $\mathbb{Z}_p$ é imagem por θ_j de p^{t-1} elementos de $\mathbb{Z}_p^t$, donde $\omega_X(p) = \mathbb{E}(\mathbf{1}_{\theta_j(x)\equiv 0 (\mathrm{mod}\ p)} \mid x \in \mathbb{Z}_p^t) = 1/p$.

Suponhamos agora que $p > w(N)$ e $|X| = 2$. Vamos ver que nenhuma das formas lineares puras $W(\psi_i - b_i)$ é múltiplo de nenhuma outra módulo p. De fato, se fosse o caso, teríamos $L_{ij} \equiv \lambda L_{i'j} (\mathrm{mod}\ p)$ para um certo λ e todo $j \le t$, mas, se a/q e a'/q' são dois racionais na forma simplificada com $|a|, |a'|, q, q' < \sqrt{w(N)}/2$ e $a/q = a'/q' (\mathrm{mod}\ p)$ então $a = a'$, $q = q'$. Portanto, todos os racionais $L_{ij}/L_{i'j}$, $1 \le j \le t$ são iguais, e logo as formas lineares puras $\psi_i - b_i$ e $\psi_{i'} - b_{i'}$ são múltiplos racionais uma da outra, absurdo. Portanto, o conjunto dos $x \in (\mathbb{Z}/p\mathbb{Z})^t$ para os quais $\theta_i(x) \equiv 0 (\mathrm{mod}\ p)$ para todo $i \in X$ está contido na interseção de dois subespaços afins de $(\mathbb{Z}/p\mathbb{Z})^t$, e portanto tem no máximo p^{t-2} elementos, donde $\omega_X(p) \le p^{-2}$. □

O lema acima implica, comparando com a definição de $E_p(z, z')$ que

$$E_p(z,z') = 1 - \mathbf{1}_{p>w(N)} \sum_{j=1}^{m} \left(p^{-1-z_j} + p^{-1-z'_j} - p^{-1-z_j-z'_j}\right) + \quad (3.4)$$

$$+ \mathbf{1}_{p>w(N)} \sum_{\substack{X,X' \subset \{1,\dots,m\} \\ |X \cup X'| \ge 2}} \frac{\mathcal{O}(1/p^2)}{p^{\sum_{j\in X} z_j + \sum_{j \in X'} z'_j}},$$

onde o numerador $\mathcal{O}(1/p^2)$ não depende de z, z'.

Vamos agora fatorar E_p como $E_p = E_p^{(1)} E_p^{(2)} E_p^{(3)}$, onde

$$E_p^{(1)}(z,z') := \frac{E_p(z,z')}{\prod_{j=1}^{m} \left(1 - \mathbf{1}_{p>w(N)} p^{-1-z_j}\right)\left(1 - \mathbf{1}_{p>w(N)} p^{-1-z'_j}\right)\left(1 - \mathbf{1}_{p>w(N)} p^{-1-z_j-z'_j}\right)^{-1}}$$

$$E_p^{(2)}(z,z') := \prod_{j=1}^{m} \left(1 - \mathbf{1}_{p\le w(N)} p^{-1-z_j}\right)^{-1} \quad \left(1 - \mathbf{1}_{p\le w(N)} p^{-1-z'_j}\right)^{-1} \times \\ \times \left(1 - \mathbf{1}_{p\le w(N)} p^{-1-z_j-z'_j}\right)$$

$$E_p^{(3)}(z,z') := \prod_{j=1}^{m} \left(1 - p^{-1-z_j}\right)\left(1 - p^{-1-z'_j}\right)\left(1 - p^{-1-z_j-z'_j}\right)^{-1}.$$

Definindo $G_j := \prod_{p \text{ primo}} E_p^{(j)}$ para $j = 1, 2, 3$ temos $F = G_1 G_2 G_3$ (pelo menos para $\mathrm{Re}(z_j)$, $\mathrm{Re}(z'_j)$ suficientemente grandes). Em termos da função ζ de Riemann, $\zeta(s) = \prod_{p \text{ primo}} \left(1 - 1/p^s\right)^{-1}$, temos $G_3(z,z') = \prod_{j=1}^{m} \dfrac{\zeta(1+z_j+z'_j)}{\zeta(1+z_j)\zeta(1+z'_j)}$, e em particular G_3 é holomorfa em $(\mathrm{Re}\ z > 0)^{2m}$, e se estende de forma meromofa a uma vizinhança do fecho deste domínio (na verdade a todo o $\mathbb{C}^{2m}$).

Para os outros fatores, faremos estimativas que permitem continuá-los analiticamente um pouco à esquerda dos eixos imaginários.

Definição 3.2.1. *Para cada $\sigma > 0$, seja $D_\sigma^m \subset \mathbb{C}^{2m}$ o domínio*

$$D_\sigma^m = \left\{ z_j, z'_j \mid -\sigma < Re(z_j), Re(z'_j) < 100, 1 \le j \le m \right\}.$$

Se $G = G(z,z')$ é uma função analítica de $2m$ variáveis complexas em D_σ^m, definimos a norma de G em $C^k(D_\sigma^m)$ para cada $k \in \mathbb{N}$ como

$$||G||_{C^k(D_\sigma^m)} = \sup_{a_1+\cdots+a_m+a'_1+\cdots+a'_m \le k} \\ \left\| \left(\frac{\partial}{\partial z_1}\right)^{a_1} \cdots \left(\frac{\partial}{\partial z_m}\right)^{a_m} \left(\frac{\partial}{\partial z'_1}\right)^{a'_1} \cdots \left(\frac{\partial}{\partial z'_m}\right)^{a'_m} G \right\|_{L^\infty(D_\sigma^m)}$$

onde $a_1, \ldots, a_m, a'_1, \ldots, a'_m$ percorrem os inteiros não negativos com soma menor ou igual a k.

Lema 3.2.3. *Os produtos de Euler* $\prod_{p \text{ primo}} E_p^{(j)}$ *para* $j = 1, 2$ *são absolutamente convergentes no domínio* $D^m_{1/30m}$. *Em particular,* G_1, G_2 *podem ser continuadas analiticamente a esse domínio. Além disso, temos as estimativas*

$$||G_1||_{C^m(D^m_{1/30m})} \leq \mathcal{O}_m(1),$$
$$||G_2||_{C^m(D^m_{1/30m})} \leq \mathcal{O}_{m,w(N)}(1),$$
$$G_1(0,0) = 1 + o_m(1) \quad e \quad G_2(0,0) = (W/\phi(W))^m.$$

Nota: Os resultados do Capítulo 1 sobre a função ζ mostram que G_3 se estende meromorficamente a $D^m_{1/2} \supset D^m_{1/30m}$. A escolha de $\sigma = 1/30m$ no lema acima não é a melhor possível, mas qualquer σ positivo dependendo só de m seria suficiente. A dependência do termo $\mathcal{O}_{m,w(N)}(1)$ em $w(N)$ não é importante, mas é possível obter sem muita dificuldade cotas do tipo $w(N)^{\mathcal{O}_m(w(N))}$.

Demonstração. Vamos considerar inicialmente o caso $j = 1$. De (3.4) e da expansão em série de Taylor das funções envolvidas, temos a estimativa grosseira $E_p^{(1)}(z, z') = 1 + \mathcal{O}_m\big(p^{-2+2/30m}\big)$ em $D^m_{1/30m}$, o que dá a convergência do produto e a estimativa de G_1 em $C^m(D^m_{1/30m})$. A estimativa para $G_1(0,0)$ também segue daí, pois os fatores do produto são identicamente iguais a 1 para $p \leq w(N)$.

A estimativa para G_2 é fácil pois G_2 é um produto finito de no máximo $w(N)$ termos, e a fórmula para $G_2(0,0)$ segue diretamente de $\prod_{\substack{p \text{ primo} \\ p \leq w(N)}} \left(\frac{p-1}{p}\right) = \frac{\phi(W)}{W}$. □

Para estimar o lado esquerdo de (3.2), que escrevemos sob forma de integral, precisamos do seguinte lema devido a Goldston e Yıldırım, que provaremos posteriormente, o qual estima integrais de contorno como as que aparecem nesse contexto:

Lema 3.2.4. *Seja* R *um real positivo e seja* $G = G(z, z')$ *uma função analítica em* $2m$ *variáveis complexas no domínio* D^m_σ *para algum* $\sigma >$

0. *Suponha que*

$$||G||_{C^m(D_\sigma^m)} = \exp\big(\mathcal{O}_{m,\sigma}((\log R)^{1/15})\big).$$

Então

$$\begin{aligned}
&\frac{1}{(2\pi i)^{2m}} \int_{\Gamma_1} \dots \int_{\Gamma_1} G(z,z') \prod_{j=1}^{m} \frac{\zeta(1+z_j+z'_j)}{\zeta(1+z_j)\zeta(1+z'_j)} \frac{R^{z_j+z'_j}}{z_j^2 z_j'^2}\, dz_j\, dz'_j \\
&= G(0,\dots,0)(\log R)^m + \sum_{j=1}^{m} \mathcal{O}_{m,\sigma}\big(||G||_{C^j(D_\sigma^m)}(\log R)^{m-j}\big) \\
&\quad + \mathcal{O}_{m,\sigma}\big(e^{-\delta(\log R)^{1/10}}\big),
\end{aligned}$$

para um certo $\delta = \delta(m) > 0$.

Vamos aplicar este lema com $G = G_1 G_2$ e $\sigma = 1/30m$. Pelo Lema 3.2.3 e pela regra de Leibnitz, obtemos as estimativas

$$||G||_{C^j(D^m_{1/30m})} \leq \mathcal{O}_{j,m,w(N)}(1), \text{ para todo } j \leq m.$$

Em particular, obtemos $||G||_{C^m(D_\sigma^m)} = \exp\big(\mathcal{O}_{m,\sigma}(\log R)^{1/15}\big)$ desde que o crescimento de $w(N)$ seja suficientemente lento. O Lema 3.2.3 também nos dá $G(0,0) = (1+o_m(1))(W/\phi(W))^m$. Concluímos que, se o crescimento de $w(N)$ é suficientemente lento, nossa expressão integral para o lado esquerdo de (3.2) é, pelo Lema 3.2.4, $(1+o_m(1)\big(W\log R/\phi(W)\big)^m$, o que conclui a prova da Proposição 3.1.1.

3.3 Correlações de ordem superior de Λ_R

Vamos agora adaptar os argumentos acima para provar a Proposição 3.1.2. Temos agora apenas uma variável, mas não podemos usar o Lema 3.2.2, pois as formas lineares só podem diferir pelos termos constantes nesse caso. Contudo, os argumentos anteriores a este lema continuam funcionando. Em particular, podemos escrever o lado esquerdo da desigualdade do enunciado da Proposição 3.1.2 como

$$(2\pi i)^{-2m} \int_{\Gamma_1} \dots \int_{\Gamma_1} F(z,z') \prod_{j=1}^{m} \frac{R^{z_j+z'_j}}{z_j^2 z_j'^2}\, dz_j\, dz'_j\,,$$

onde F é definido como em (3.3), com a diferença de que agora $\omega_X(p)$ deve ser definido como

$$\omega_X(p) := \mathbb{E}\big(\prod_{i\in X} \mathbf{1}_{W(x+h_i)+1\equiv 0(\mathrm{mod}\ p)} \mid x \in \mathbb{Z}_p\big).$$

Temos ainda $\omega_\phi(p) = 1$ para todo p. O análogo do Lema 3.2.2 é o seguinte:

Lema 3.3.1. *Se $p \le w(N)$, então $\omega_X(p) = 0$ para todo $X \neq \varnothing$. Em particular, $E_p = 1$ quando $p \le w(N)$. Se $p > w(N)$, então $\omega_X(p) = p^{-1}$ quando $|X| = 1$ e $\omega_X(p) \le p^{-1}$ quando $|X| \ge 2$. Além disso, quando $|X| \ge 2$, temos $\omega_X(p) = 0$ sempre que p não divide $\Delta := \prod_{1\le i<j\le s} |h_i - h_j|$.*

Demonstração. Quando $p \le w(N)$, temos $W(x+h_i)+1 \equiv 1 \pmod p$, donde segue nossa afirmação. Quando $p > w(N)$, e $|X| \ge 1$, $\omega_X(p) = 1/p$ quando todas as classes de congruência $h_i (\mathrm{mod}\ p)$, $i \in X$ são iguais, e $\omega_X(p) = 0$ caso contrário, e daí segue o resultado. □

Aplicando o lema acima, obtemos o seguinte análogo de (3.4):

$$E_p(z,z') = 1 - \mathbf{1}_{p>w(N)} \sum_{j=1}^{m} \big(p^{-1-z_j} + p^{-1-z'_j} - p^{-1-z_j-z'_j}\big) + \mathbf{1}_{p>w(N),p|\Delta}\, \lambda_p(z,z')$$

onde $\lambda_p(z,z')$ é uma expressão da forma

$$\lambda_p(z,z') = \sum_{\substack{X,X'\subset\{1,\dots,m\}\\ |X\cup X'|\ge 2}} \frac{\mathcal{O}(1/p)}{p^{\sum_{j\in X} z_j + \sum_{j\in X'} z'_j}}$$

na qual a quantidade $\mathcal{O}(1/p)$ não depende de z, z'. Podemos então

fatorar $E_p = E_p^{(0)} E_p^{(1)} E_p^{(2)} E_p^{(3)}$, onde

$$E_p^{(0)} = 1 + \mathbf{1}_{p>w(N),p|\Delta} \cdot \lambda_p(z, z')$$

$$E_p^{(1)} = \frac{E_p}{E_p^{(0)} \prod_{j=1}^{m} (1 - \mathbf{1}_{p>w(N)} p^{-1-z_j})(1 - \mathbf{1}_{p>w(N)} p^{-1-z'_j})(1 - \mathbf{1}_{p>w(N)} p^{-1-z_j-z'_j})^{-1}}$$

$$E_p^{(2)} = \prod_{j=1}^{m} (1 - \mathbf{1}_{p\leq w(N)} p^{-1-z_j})^{-1} (1 - \mathbf{1}_{p\leq w(N)} p^{-1-z'_j})^{-1} (1 - \mathbf{1}_{p\leq W(N)} p^{-1-z_j-z'_j})$$

$$E_p^{(3)} = \prod_{j=1}^{m} (1 - p^{-1-z_j})(1 - p^{-1-z'_j})(1 - p^{-1-z_j-z'_j})^{-1}.$$

Seja $G_j = \prod_{p \text{ primo}} E_p^{(j)}$. Então, como antes, $F = G_0 G_1 G_2 G_3$, e G_3 é dado por $\prod_{j=1}^{m} \frac{\zeta(1+z_j+z'_j)}{\zeta(1+z_j)\zeta(1+z'_j)}$, como antes. Para G_0, G_1 e G_2, temos o seguinte análogo ao Lema 3.2.3:

Lema 3.3.2. *Seja $0 < \sigma \leq 1/30m$. Os produtos de Euler $\prod_{p \text{ primo}} E_p^{(\ell)}$ para $\ell = 0, 1, 2$ são absolutamente convergentes no domínio D_σ^m. Em particular, G_0, G_1 e G_2 podem ser continuados analiticamente a esse domínio. Além disso, temos as estimativas seguintes:*

$$||G_0||_{C^r(D_\sigma^m)} \leq \mathcal{O}_m(1) \cdot \big(\frac{\log R}{\log\log R}\big)^r \prod_{\substack{p|\Delta \\ p \text{ primo}}} \big(1 + \mathcal{O}_m(p^{2m\sigma-1})\big)$$

$$||G_0||_{C^m(D_\sigma^m)} \leq \exp\big(\mathcal{O}_m((\log R)^{1/15})\big)$$

$$||G_1||_{C^m(D_\sigma^m)} \leq \mathcal{O}_m(1)$$

$$||G_2||_{C^m(D_\sigma^m)} \leq \mathcal{O}_{m,w(N)}(1)$$

$$G_0(0,0) = \prod_{\substack{p|\Delta \\ p \text{ primo}}} \big(1 + \mathcal{O}_m(p^{-1/2})\big)$$

$$G_1(0,0) = 1 + \mathcal{O}_m(1)$$

$$G_2(0,0) = (W/\phi(W))^m.$$

Demonstração. As estimativas para G_1 e G_2 podem ser provadas exatamente como no Lema 3.2.3 (os fatores adicionais $\lambda_p(z,z')$ que aparecem no numerador e no denominador de $E_p^{(1)}$ se cancelam em primeira ordem, e portanto não criam dificuldades adicionais); vamos portanto nos dedicar às estimativas sobre G_0.

Temos $G_0 = \prod_{\substack{p|\Delta \\ p \text{ primo}}} E_p^{(0)}$. O número de primos que dividem Δ é no máximo $\mathcal{O}(\log\Delta/\log\log\Delta)$ (ver apêndice). Usando a estimativa grosseira

$$\Delta = \prod_{1\le i<j\le m} |h_i - h_j| \le N^{m^2} \le R^{\mathcal{O}_m(1)},$$

vemos que o número de fatores no produto de Euler é $\mathcal{O}(\log R/\log\log R)$. Diferenciando r vezes para $0 \le r \le m$ por meio da regra de Leibnitz, obtemos uma soma de $\mathcal{O}_m((\log R/\log\log R)^r)$ termos, cada um dos quais consistindo de $\mathcal{O}_m(\log R/\log\log R)$ fatores, os quais são iguais a alguma derivada de $1+\lambda_p(z,z')$, de alguma ordem entre 0 e r. Em D_σ^m, cada fator é limitado por $1+\mathcal{O}_m(p^{2m\sigma-1})$ (na verdade, os termos que contêm um número positivo de derivadas serão muito menores, pois o termo constante 1 é eliminado). Isso nos dá a primeira estimativa sobre $||G_0||_{C^r(D_\sigma^m)}$.
Para provar a estimativa seguinte, basta mostrar que

$$\prod_{\substack{p|\Delta \\ p \text{ primo}}} \big(1+\mathcal{O}_m(p^{2m\sigma-1})\big) \le \exp\big(\mathcal{O}_m((\log R)^{1/15})\big).$$

Tomando logaritmos e usando a hipótese $\sigma \le 1/30m$, é suficiente provar que $\sum_{p|\Delta} p^{-14/15} \le \mathcal{O}((\log\Delta)^{1/15})$, pois $\Delta \le R^{\mathcal{O}_m(1)}$. Para isso, como Δ tem no máximo $\mathcal{O}(\log\Delta/\log\log\Delta)$ fatores primos (ver apêndice), temos

$$\sum_{p|\Delta} p^{-14/15} \le \sum_{1\le n\le \mathcal{O}(\log\Delta/\log\log\Delta)} n^{-14/15} = \mathcal{O}((\log\Delta)^{1/15}),$$

como queríamos.

A estimativa para $G_0(0,0)$ segue da estimativa grosseira $E_p^{(0)}(z,z') = 1+\mathcal{O}_m(p^{-1/2})$. □

Aplicamos agora o Lema 3.2.4 com $\sigma := 1/30m$ e $G := G_0G_1G_2$. Ainda pela regra de Leibnitz, temos

$$||G||_{C^m(D_\sigma^m)} = \exp\big(\mathcal{O}_{m,\sigma}((\log R)^{1/15})\big),$$

donde, pelo lema,

$$\frac{1}{2\pi i}\int_{\Gamma_1}\dots\int_{\Gamma_1} F(z,z')\prod_{j=1}^{m}\frac{R^{z_j+z_j'}}{z_j^2 z_j'^2}\,dz_j\,dz_j' \leq$$
$$\leq \mathcal{O}_m(1)\big(\frac{W}{\phi(W)}\big)^m(\log R)^m\prod_{p|\Delta}\big(1+\mathcal{O}_m(p^{-1/2})\big)+$$
$$+\,\mathcal{O}_{m,w(N)}\big(\frac{(\log R)^m}{\log\log R}\big)\prod_{p|\Delta}\big(1+\mathcal{O}_m(p^{-1/2})\big)+\mathcal{O}_m\big(e^{-\delta(\log R)^{1/10}}\big),$$

e, escolhendo $w(N)$ que cresça de modo suficientemente lento em relação a N (e logo também em relação a R), o primeiro termo dominará os demais, o que conclui a prova da Proposição 3.1.2.

Nota: De fato o argumento acima pode ser usado para dar uma estimativa assintótica para o lado esquerdo da desigualdade no enunciado da Proposição 3.1.2, em vez de fornecer apenas uma cota superior. Para isso, basta estimar $G_0(0,0)$ mais cuidadosamente. Isto foi feito em detalhes por Goldston e Yıldırım no caso $W = 1$.

3.4 Prova do Lema 3.2.4

Provaremos agora o Lema 3.2.4. No que segue, $R \geq 2$, $m \geq 1$ e $\sigma > 0$ serão fixados. Usaremos $\delta > 0$ para denotar diversas constantes pequenas, que podem variar de acordo com as retas verticais onde faremos integração.

Vamos recordar a região livre de zeros para a função ζ de Riemann obtida no (apêndice ao) Capítulo 1:

$$Z := \big\{z \in \mathbb{C} \mid 10 \geq \text{Re } z \geq 1 - \frac{\beta}{(\log(|\text{Im } z|+2))^9}\big\},$$

para um certo $\beta \in (0,1)$ pequeno é uma região tal que ζ é não-nula e meromorfa em Z com um único pólo simples em 1. Além disso, temos as seguintes estimativas, válidas para todo $s \in Z$:

$$\zeta(s) - \frac{1}{s-1} = \mathcal{O}(|\text{Im } s| + 2); \quad \frac{1}{\zeta(s)} = \mathcal{O}((|\text{Im } s| + 2)^7).$$

Temos ainda que, se Re $s \geq 3/4$, então $\zeta(s) - \frac{1}{s-1} = \mathcal{O}((|\text{Im } s| + 2)^{1/4})$.

Podemos escolher β pequeno de modo que Z está contido na região onde $\max\{1-\sigma, 7/8\} < \text{Re } s < 101$. As constantes envolvidas na notação $\mathcal{O}(\quad)$ podem depender de β e σ, sem necessidade de menção explícita.

Além do caminho Γ_1 dado por $\Gamma_1(t) = 1/\log R + it$, $-\infty < t < \infty$, definiremos dois outros caminhos:

$$\Gamma_0(t) := \frac{\beta}{(\log(|t|+2))^9} + it, \; -\infty < t < \infty$$
$$\text{e} \quad \Gamma_2(t) := 1 + it, \; -\infty < t < \infty.$$

Assim, Γ_0 é a fronteira esquerda de $Z - 1$, situada à esquerda da origem, enquanto Γ_1 e Γ_2 estão situadas à direita da origem. A utilidade de Γ_2 vem do fato de que $\zeta(1 + z + z')$ não tem nenhum pólo quando $z \in Z - 1$ e $z' \in \Gamma_2$ (mas não estimaremos integrais em Γ_2).

O próximo lema fornece estimativas para as integrais seguintes:

Lema 3.4.1. *Seja B uma constante fixada. Temos as seguintes estimativas:*

$$\int_{\Gamma_0} \Big|\frac{R^z\,dz}{z^{3/2}}\Big| \leq \mathcal{O}\big(e^{-\delta(\log R)^{1/10}}\big)$$
$$\int_{\Gamma_1} (\log(|z|+2))^B \Big|\frac{R^z\,dz}{z^2}\Big| \leq \mathcal{O}_B(\log R)$$

onde $\delta = \delta(\beta) > 0$ é uma constante independente de R.

Demonstração. Como $|\Gamma_0'(t)| = \mathcal{O}(1)$ e $|z| \geq c(|t| + \beta)$ em Γ_0 para

uma certa constante $c > 0$, temos, para cada $T \geq 2$,

$$\int_{\Gamma_0} \big|\frac{R^z dz}{z^{3/2}}\big| \leq \mathcal{O}(1) \int_0^\infty \frac{R^{-\beta/(\log(|t|+2))^9}}{(|t|+\beta)^{3/2}}\, dt \leq$$
$$\leq \mathcal{O}(1)\big(\int_0^T R^{-\beta/(\log(|t|+2))^9}\, dt + \int_T^\infty \frac{dt}{t^{3/2}}\big) \leq$$
$$\leq \mathcal{O}(1)\big(T\, \exp(-\beta \log R/2(\log T)^9\big) + T^{-1/2}\big).$$

Escolhendo $T = \exp((\beta \log R/3)^{1/10})$, os dois termos da soma são iguais, e da ordem de $\mathcal{O}(1)\exp(-\frac{1}{2}(\beta \log R/3)^{1/10}) = \mathcal{O}\big(e^{-\delta(\log R)^{1/10}}\big)$, o que demonstra a primeira estimativa do lema.

Para a segunda estimativa, basta usar o fato de R^z ser limitado em Γ_1, donde, dividindo o intervalo de parâmetros em $\{|t| \leq 1/\log R\}$ e $\{|t| > 1/\log R\}$, obtemos a estimativa desejada, pois $\{|t| \leq 1/\log R\}$ é um intervalo de tamanho $2/\log R$ onde o integrando tem módulo $\mathcal{O}((\log R)^2)$, enquanto

$$\int_{1/\log R}^\infty \big(\log(t+2)\big)^B \cdot \frac{dt}{t^2}$$
$$= \int_{1/\log R}^1 \big(\log(t+2)\big)^B \cdot \frac{dt}{t^2}$$
$$+ \int_1^\infty \big(\log(t+2)\big)^B \cdot \frac{dt}{t^2}$$
$$= \mathcal{O}_B\big(\int_{1/\log R}^1 \frac{dt}{t^2}\big) + \mathcal{O}_B(1)$$
$$= \mathcal{O}_B(\log R) + \mathcal{O}_B(1) = \mathcal{O}_B(\log R).$$

□

O próximo lema está relacionado com o caso $m = 1$ do Lema 3.2.4:

Lema 3.4.2. *Seja $f(z, z')$ analítica em D_σ^1 e suponha que*

$$|f(z, z')| \leq \exp\big(\mathcal{O}_m(\log R)^{1/15}\big)$$

uniformemente nesse domínio. Então a integral

$$I := \frac{1}{(2\pi i)^2}\int_{\Gamma_1}\int_{\Gamma_1} f(z,z')\,\frac{\zeta(1+z+z')}{\zeta(1+z)\zeta(1+z')}\,\frac{R^{z+z'}}{z^2 z'^2}\,dzdz'$$

satisfaz a estimativa

$$I = f(0,0)\log R + \frac{\partial f}{\partial z'}(0,0) + \frac{1}{2\pi i}\int_{\Gamma_0} f(z,-z)\,\frac{dz}{\zeta(1+z)\zeta(1-z)z^4} + \mathcal{O}_m\big(e^{-\delta(\log R)^{1/10}}\big),$$

para um certo $\delta = \delta(\sigma,\beta) > 0$ independente de R.

Demonstração. Observamos que temos decaimento suficiente no integrando para trocar a ordem de integração, e para mover caminhos de integração em cada variável z, z' mantendo a outra fixa, sem dificuldades quando $\mathrm{Im}(z)$, $\mathrm{Im}(z') \to \infty$, pelas estimativas sobre ζ na região livre de zeros Z. Devemos apenas levar em conta o efeito de mover caminhos de integração através de um pólo do integrando. Em particular podemos mover o caminho de z' de Γ_1 para Γ_2, pois não passamos por nenhum pólo do integrando nesse processo. Consideremos o integrando para cada $z' \in \Gamma_2$ como uma função analítica de z, e vamos tentar mover o caminho de integração em z para Γ_0. Nesse processo passamos por um único pólo em $z = 0$. O resíduo nesse pólo é $\dfrac{1}{(2\pi i)^2}\displaystyle\int_{\Gamma_2} f(0,z')\,\frac{R^{z'}}{z'^2}\,dz'$, e portanto temos $I = I_1 + I_2$, onde

$$I_1 := \frac{1}{2\pi i}\int_{\Gamma_2} f(0,z')\,\frac{R^{z'}}{z'^2}\,dz' \quad \text{e}$$

$$I_2 := \frac{1}{(2\pi i)^2}\int_{\Gamma_2}\int_{\Gamma_0} f(z,z')\,\frac{\zeta(1+z+z')R^{z+z'}}{\zeta(1+z)\zeta(1+z')z^2 z'^2}\,dzdz'.$$

Para estimar I_1, movemos o caminho de integração para Γ_0. Como antes, há apenas um pólo, duplo, em $z' = 0$. O resíduo nesse pólo é $\dfrac{1}{2\pi i}\big(f(0,0)\log R + \dfrac{\partial f}{\partial z'}(0,0)\big)$, e portanto

$$I_1 = f(0,0)\log R + \frac{\partial f}{\partial z'}(0,0) + \frac{1}{2\pi i}\int_{\Gamma_0} f(0,z')\,\frac{R^{z'}}{z'^2}\,dz' =$$

$$= f(0,0)\log R + \frac{\partial f}{\partial z'}(0,0) + \mathcal{O}_m\big(e^{-\delta(\log R)^{1/10}}\big)$$

para um certo $\delta > 0$. A última igualdade é conseqüência da primeira estimativa para f no Lema 3.4.1 (com $B = 0$).

Para estimar I_2 , olhamos o integrando como uma função analítica de z'. Movemos o caminho de integração em z' de Γ_2 para Γ_0 , o que está autorizado pelo decaimento em faixas verticais quando $|\text{Im } z'| \to \infty$ do integrando. Fazendo isso, atravessamos exatamente dois pólos simples, em $z' = -z$ e em $z' = 0$. O resíduo no primeiro é

$$\frac{1}{(2\pi i)^2} \int_{\Gamma_0} f(z, -z)\, \frac{dz}{\zeta(1+z)\zeta(1-z)z^4}\,,$$

o que fornece um dos termos em nossa fórmula para I.

O resíduo em $z' = 0$ é $\dfrac{1}{(2\pi i)^2} \displaystyle\int_{\Gamma_0} f(z, 0)\, \frac{R^z}{z^2}\, dz$, que é

$$\mathcal{O}\big(e^{-\delta(\log R)^{1/10}}\big) \cdot \exp\big(\mathcal{O}((\log R)^{1/15})\big) = \mathcal{O}\big(e^{-\tilde{\delta}(\log R)^{1/10}}\big),$$

com $\tilde{\delta} = \delta/2$.

O valor de I_2 é a soma dessas duas quantidades com a integral sobre o novo caminho de integração Γ_0 , que é

$$\int_{\Gamma_0} \int_{\Gamma_0} f(z, z')\, \frac{\zeta(1+z+z')R^{z+z'}}{\zeta(1+z)\zeta(1+z')z^2 z'^2}\, dz dz'.$$

Nesse integrando temos $|f| = \exp(\mathcal{O}_m((\log R)^{1/15}))$. Por outro lado, temos

$$\Big|\frac{1}{\zeta(1+z)}\Big| = \mathcal{O}((\log(|\text{Im } z|+2))^7), \Big|\frac{1}{\zeta(1+z')}\Big| = \mathcal{O}((\log(|\text{Im } z'|+2))^7)$$

e, como Re z, Re $z' \geq -1/8$ em Γ_0 , Re $(z + z') \geq 3/4$, donde $|\zeta(1 + z + z')| = \mathcal{O}((|\text{Im}(z + z')| + 2)^{1/4})$, e portanto

$$\begin{aligned}
&\Big|\frac{\zeta(1+z+z')}{\zeta(1+z)\zeta(1+z')}\Big| = \\
&= \mathcal{O}((\log(|\text{Im } z| + 2))^7)(\log(|\text{Im } z'| + 2))^7) \times \\
&\quad \times (|\text{Im } z| + 2)^{1/4}(|\text{Im } z'| + 2)^{1/4}) = \\
&= \mathcal{O}((|\text{Im } z| + 2)^{1/2}(|\text{Im } z'| + 2)^{1/2}) = \mathcal{O}(|z|^{1/2}|z'|^{1/2}),
\end{aligned}$$

pelas estimativas sobre a função ζ descritas anteriormente. Assim, usando duas vezes (para z e z') a primeira estimativa do Lema 3.4.1, obtemos que a integral em questão é $\mathcal{O}_m\big(e^{-\delta(\log R)^{1/10}}\big)$, para um certo $\delta > 0$.

Obtemos assim estimativas para I_1 e I_2 com erros que são $\mathcal{O}\big(e^{-\delta(\log R)^{1/10}}\big)$. Somando essas estimativas, completamos a prova do lema. □

Prova do Lema 3.2.4. Seja $G = G(z, z')$ uma função analítica de $2m$ variáveis complexas no domínio D_σ^m satisfazendo a hipótese

$$||G||_{C^m(D_\sigma^m)} = \exp(\mathcal{O}_{m,\sigma}((\log R)^{1/15})).$$

No que segue permitiremos que as constantes implícitas no notação $\mathcal{O}(\quad)$ dependam de m, β, σ. Queremos provar que a integral

$$I(G,m) := \frac{1}{(2\pi i)^{2m}} \int_{\Gamma_1} \dots \int_{\Gamma_1} G(z,z') \times \\ \times \prod_{j=1}^{m} \frac{\zeta(1+z_j+z_j')}{\zeta(1+z_j)\zeta(1+z_j')} \frac{R^{z_j+z_j'}}{z_j^2 z_j'^2}\, dz_j dz_j' \,,$$

satisfaz a estimativa

$$I(G,m) = G(0,\dots,0)(\log R)^m + \sum_{i=1}^{m} \mathcal{O}(||G||_{C^j(D_\sigma^m)}(\log R)^{m-j}) \\ + \mathcal{O}(e^{-\delta(\log R)^{1/10}}).$$

A prova é por indução em m. O caso $m = 1$ segue do Lema 3.4.2, pois

$$\Big|\frac{\partial f}{\partial z_1'}(0,0)\Big| = \mathcal{O}(||G||_{C^1(D_\sigma^1)})$$

e

$$\Big|\frac{1}{2\pi i}\int_{\Gamma_0} G(z_1,-z_1)\,\frac{dz_1}{\zeta(1+z_1)\zeta(1-z_1)z_1^4}\Big| \\ = \mathcal{O}(||G||_{C^0(D_\sigma^1)}) = \mathcal{O}(||G||_{C^1(D_\sigma^1)}),$$

o que, por sua vez, segue de

$$\int_{\Gamma_0} \Big| \frac{dz_1}{\zeta(1+z_1)\zeta(1-z_1)z_1^4} \Big| = \mathcal{O}(1).$$

A última estimativa é uma conseqüência simples de nossas estimativas para ζ em Z.

Suponhamos agora que vale o resultado para um certo $m \geq 1$. Queremos prová-lo para $m+1$. Aplicando o Lema 3.4.2 nas variáveis z_{m+1}, z'_{m+1}, obtemos

$$\begin{aligned}
&I(G, m+1) = \\
&= \frac{\log R}{(2\pi i)^{2m}} \int_{\Gamma_1} \dots \int_{\Gamma_1} G(z_1, \dots, z_m, 0, z'_1, \dots, z'_m, 0) \times \\
&\quad \times \prod_{j=1}^{m} \frac{\zeta(1+z_j+z'_j)}{\zeta(1+z_j)\zeta(1+z'_j)} \frac{R^{z_j+z'_j}}{z_j^2 z_j'^2} dz_j dz'_j + \\
&\quad + \frac{1}{(2\pi i)^{2m}} \int_{\Gamma_1} \dots \int_{\Gamma_1} (H(z_1, \dots, z_m, z'_1, \dots, z'_m) \\
&\quad + r(z_1, \dots, z_m, z'_1, \dots, z'_m)) \prod_{j=1}^{m} \frac{\zeta(1+z_j+z'_j)}{\zeta(1+z_j)\zeta(1+z'_j)} \frac{R^{z_j+z'_j}}{z_j^2 z_j'^2} dz_j dz'_j = \\
&= I(G(z_1, \dots, z_m, 0, z'_1, \dots, z'_m, 0), m) \log R + I(H, m),
\end{aligned}$$

onde $\delta > 0$, $H \colon D_\sigma^m \to \mathbb{C}$ é a função definida por

$$\begin{aligned}
&H(z_1, \dots, z_m, z'_1, \dots, z'_m) := \frac{\partial G}{\partial z'_{m+1}}(z_1, \dots, z_m, 0, z'_1, \dots, z'_m, 0) + \\
&+ \frac{1}{2\pi i} \int_{\Gamma_0} G(z_1, \dots, z_m, z_{m+1}, z'_1, \dots, z'_m, \dots, -z_{m+1}) \times \\
&\times \frac{dz_{m+1}}{\zeta(1+z_{m+1})\zeta(1-z_{m+1})z_{m+1}^4}
\end{aligned}$$

e

$$\begin{aligned}
&r(z_1,\dots,z_m,z'_1,\dots,z'_m)\\
&:=\frac{1}{(2\pi i)^2}\int_{\Gamma_1}\int_{\Gamma_1} G(z_1,\dots,z_{m+1},z'_1,\dots,z'_{m+1})\times\\
&\times\frac{\zeta(1+z_{m+1}+z'_{m+1})}{\zeta(1+z_{m+1})\zeta(1+z'_{m+1})}\frac{R^{z_{m+1}+z'_{m+1}}}{z_{m+1}^2 z_{m+1}'^2}dz_{m+1}dz'_{m+1}\\
&\quad -H(z_1,\dots,z_m,z'_1,\dots,z'_m).
\end{aligned}$$

As funções $G(z_1,\dots,z_m,0,z'_1,\dots,z'_m,0)$ e $H(z_1,\dots,z_m,z'_1,\dots,z'_m)$ são analíticas em D_σ^m e, como $\int_{\Gamma_0}|\frac{dz}{\zeta(1+z)\zeta(1-z)z^4}|=\mathcal{O}(1)$, temos

$$||H||_{C^j(D_\sigma^m)}=\mathcal{O}_m(||G||_{C^{j+1}(D_\sigma^{m+1})}),\ \text{para } 0\le j\le m,$$

donde (usando também o caso $m=1$) $||r||_{C^j(D_\sigma^m)}=\mathcal{O}_m(||G||_{C^{j+1}(D_\sigma^{m+1})})$. Além disso, $|r|=\mathcal{O}(e^{-\delta(\log R)^{1/10}})$, pelo Lema 3.4.2. Temos portanto, usando a hipótese de indução,

$$\begin{aligned}
I(G,m+1)&=G(0,\dots,0)(\log R)^{m+1}+\\
&+\sum_{j=1}^m\mathcal{O}_m(||G(\cdot,0,\cdot,0)||_{C^j(D_\sigma^m)}(\log R)^{m+1-j})+\\
&+H(0,\dots,0)(\log R)^m+\sum_{j=1}^m\mathcal{O}_m(||H||_{C^j(D_\sigma^m)}(\log R)^{m-j})+\\
&+r(0,\dots,0)(\log R)^m+\sum_{j=1}^m\mathcal{O}_m(||r||_{C^j(D_\sigma^m)}(\log R)^{m-j})=\\
&=G(0,\dots,0)(\log R)^{m+1}+\sum_{j=1}^m\mathcal{O}_m(||G||_{C^j(D_\sigma^{m+1})}(\log R)^{m+1-j})+\\
&+H(0,\dots,0)(\log R)^m+\sum_{j=1}^m\mathcal{O}_m(||G||_{C^{j+1}(D_\sigma^{m+1})}(\log R)^{m-j})+\\
&+\mathcal{O}(e^{-\delta(\log R)^{1/10}})+\sum_{j=1}^m\mathcal{O}_m(||G||_{C^{j+1}(D_\sigma^{m+1})}(\log R)^{m-j})=
\end{aligned}$$

$$= G(0,\dots,0)(\log R)^{m+1} + \sum_{j=1}^{m+1} \mathcal{O}_m(||G||_{C^j(D_\sigma^{m+1})}(\log R)^{m+1-j})$$

$$+ \mathcal{O}(e^{-\delta(\log R)^{1/10}}),$$

que é o que queríamos provar. □

Comentários: Podemos evitar o uso do teorema de Dirichlet trocando $Wn+1$ por $Wn+b$ na definição de $\widetilde{\Lambda}(n)$, onde b satisfaz mdc $(b,W)=1$, $1 \le b < W$ e é tal que $\#\{1 \le n \le N \mid Wn+b \text{ é primo}\}$ é máximo, pois, de fato, só precisamos da estimativa $\sum_{\varepsilon_k N \le n \le 2\varepsilon_k N} \widetilde{\Lambda}(n) \ge c \cdot \varepsilon_k N$, para alguma constante positiva c. Esse mesmo truque deve ser usado na prova da generalização do teorema principal para a existência de progressões aritméticas arbitrariamente longas em conjuntos de primos com densidade positiva (tais conjuntos não necessariamente conterão primos congruentes a 1 módulo W). O resto do argumento não precisa de modificações substanciais.

Olhando retroativamente para a prova, vemos que o termo de erro no teorema principal não precisa ser $o(1)$, mas basta ser, por exemplo, $\frac{1}{2}\,c(k,\delta)+o(1)$, o que permite tomar $w(N)$ uma constante grande dependendo apenas de k. Isto faz com que a perda na proporção de primos devida à passagem de n para $Wn+1$ seja uniformemente limitada em N, o que permite provar que existe uma constante $\gamma(k)>0$ tal que o número de progressões aritméticas formadas por k números primos entre 1 e N é pelo menos $(\gamma(k)+o(1))\,\dfrac{N^2}{(\log N)^k}$. Por outro lado, argumentos da teoria do crivo mostram que o número de tais progressões aritméticas é $\mathcal{O}_k(N^2/(\log N)^k)$, e logo a estimativa inferior obtida difere do número correto apenas por um fator limitado.

3.5 Apêndice ao Capítulo 3: dois resultados elementares de teoria dos números.

I) A fórmula da inversão de Möbius.

Definimos a função de Möbius como a função $\mu\colon \mathbb{N}^* \to \mathbb{Z}$ dada por

$$\mu(n) = \begin{cases} 0 \text{ se existe } p \text{ primo tal que } p^2 \mid n \\ (-1)^k \text{ se } n = p_1 p_2 \dots p_k \text{, com } p_1 < p_2 < \dots < p_k \text{ primos} \end{cases}$$

Em particular, $\mu(1) = 1$ (1 é o produto de 0 fatores primos).

Obs.: A função μ é multiplicativa, i.e., $\text{mdc}(mn) = 1 \Rightarrow \mu(mn) = \mu(m) \cdot \mu(n)$.

É bastante comum associar a uma função $f\colon \mathbb{N}^* \to \mathbb{C}$ outra função $g\colon \mathbb{N}^* \to \mathbb{C}$ dada por $g(n) = \sum_{d|n} f(d)$. A fórmula da inversão de Möbius permite recuperar f a partir de g. Provaremos inicialmente o seguinte

Lema 3.5.1. $\sum_{d|n} \mu(d) = \begin{cases} 0, & \textit{se } n > 1 \\ 1, & \textit{se } n = 1 \end{cases}$

Demonstração. Temos $\sum_{d|1} \mu(d) = \mu(1) = 1$. Suponha agora que $n > 1$. Seja p um fator primo de n. Temos $X = \{d \geq 1; q^2 | m \Rightarrow q = 1$ e $d|n\} = Y \cup Z$, onde $Y = \{d \in X; p \nmid d\}$ e $Z = \{d \in X; p|d\} = \{p \cdot d, d \in Y\}$. Se $d|n$ e $\mu(d) \neq 0$ então $d \in X$. Por outro lado, se $d \in Y$, $\mu(p \cdot d) = -\mu(d)$, pois as paridades dos números de fatores primos de d e de $p \cdot d$ são distintas. Portanto,

$$\begin{aligned} \sum_{d|n} \mu(d) &= \sum_{d \in X} \mu(d) = \sum_{d \in Y} \mu(d) + \sum_{d \in Z} \mu(d) \\ &= \sum_{d \in Y} (\mu(d) + \mu(p \cdot d)) = \sum_{d \in Y} 0 = 0. \end{aligned}$$

□

Teorema A.2 (Fórmula da inversão de Möbius): Sejam $f, g\colon \mathbb{N}^* \to \mathbb{C}$ funções tais que $g(n) = \sum_{d|n} f(d)$, para todo $n \in \mathbb{N}^*$. Então temos $f(n) = \sum_{d|n} \mu\big(\frac{n}{d}\big) g(d)$, para todo $n \in \mathbb{N}^*$.

Demonstração. Queremos provar que

$$f(n) = \sum_{d|n} \mu\big(\frac{n}{d}\big) g(d) = \sum_{d|n} \mu\big(\frac{n}{d}\big) \big(\sum_{d'|d} f(d')\big),$$

mas

$$\begin{aligned}\sum_{d|n} \mu\big(\frac{n}{d}\big) \big(\sum_{d'|d} f(d')\big) &= \sum_{d'|n} f(d') \cdot \big(\sum_{d'|d|n} \mu\big(\frac{n}{d}\big)\big) \\ &= \sum_{d'|n} f(d') \big(\sum_{\tilde{d}|\frac{n}{d'}} \mu(\tilde{d})\big) = f(n),\end{aligned}$$

pois $\sum_{\tilde{d}|\frac{n}{d'}} \mu(\tilde{d}) = \begin{cases} 0, \text{ se } n/d' > 1, \text{ i.e., se } d' < n \\ 1, \text{ se } n/d' = 1, \text{ i.e., se } d' = n. \end{cases}$ □

II) A ordem máxima de $d(n)$.

Seja $d(n)$, para cada $n \in \mathbb{N}^*$, o número de divisores (positivos) de n. Temos então o seguinte

Teorema A.3: Para todo $\varepsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que $n > n_0 \Rightarrow d(n) < 2^{(1+\varepsilon)\log n/\log\log n}$.

Demonstração. Seja $n = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$, $p_1 < p_2 < \cdots < p_k$ primos a fatoração prima de n. Temos então

$$\begin{aligned} d(n) = \prod_{j=1}^{k} (1+\alpha_j) &= \prod_{p_j \le (\log n)^{1-\delta}} (1+\alpha_j) \times \\ &\times \prod_{p_j > (\log n)^{1-\delta}} (1+\alpha_j), \quad \text{onde } \delta = \varepsilon/2(1+\varepsilon). \end{aligned}$$

Para todo j, $1+\alpha_j \le 2^{\alpha_j}$, donde

$$\prod_{p_j>(\log n)^{1-\delta}} (1+\alpha_j) \le 2^{\sum_{p_j>(\log n)^{1-\delta}} \alpha_j}$$
$$\le 2^{\log n/\log((\log n)^{1-\delta})}$$
$$= 2^{\log n/(1-\delta)\log\log n}$$

(de fato, $n \ge \prod_{p_j>(\log n)^{1-\delta}} p_j^{\alpha_j} \ge ((\log n)^{1-\delta})^{\sum_{p_j>(\log n)^{1-\delta}} \alpha_j}$).

Por outro lado, para todo j, $1+\alpha_j \le 1+\dfrac{\log n}{\log 2}$, pois $n \ge p_j^{\alpha_j} \ge 2^{\alpha_j} \Rightarrow \log n \ge \alpha_j \log 2$. Assim,

$$\prod_{p_j\le(\log n)^{1-\delta}} (1+\alpha_j) \le \left(1+\frac{\log n}{\log 2}\right)^{(\log n)^{1-\delta}} = 2^{\mathcal{O}(\log\log n\cdot(\log n)^{1-\delta})}.$$

Temos então

$$d(n) = \prod_{j=1}^{k}(1+\alpha_j) \le 2^{\log n/(1-\delta)\log\log n+\mathcal{O}(\log n/\log\log n)}$$
$$< 2^{(1+\varepsilon)\log n/\log\log n}$$

se n é suficientemente grande, pois $\dfrac{1}{1-\delta} < \dfrac{1}{1-2\delta} = 1+\varepsilon$. □

Corolário. Para todo $\varepsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que $n > n_0 \Rightarrow \#\{p \text{ primo}, p|n\} < (1+\varepsilon)\log n/\log\log n$.

(De fato, temos $2^{\#\{p \text{ primo},p|n\}} \le d(n)$, para todo inteiro positivo n.)

Nota: É possível provar que, se $p_1 < p_2 < \cdots < p_k$ são os k primeiros números primos e $N_k = \prod_{j=1}^{k} p_j$ então $d(N_k) > 2^{\frac{\log N_k}{\log\log N_k}}$, para k suficientemente grande. De fato, $p_r < 2r\log r$, para todo r grande, donde $\log N_k = \sum_{j=1}^{k} \log p_j < \mathcal{O}(1) + \sum_{j=2}^{k}(\log 2 + \log j + \log\log j) =$

$\mathcal{O}(1) + k\log 2 + k\log k - k + o(k) + k\log\log k + o(k) = k(\log k + \log\log k - (1 - \log 2)) + o(k)$, donde

$$\begin{aligned}\log N_k/\log\log N_k &< k(\log k + \log\log k)/\log(k(\log k + \log\log k)) \\ &< k = \log d(N_k)/\log 2,\end{aligned}$$

para todo k suficientemente grande.

Referências Bibliográficas

[1] Aarão, J. O teorema dos números primos. *Tese de Mestrado.* IMPA. 1988.

[2] Brun, V. La serie $1/5 + 1/7 + 1/11 + 1/13 + 1/17 + 1/19 + 1/29 + 1/31 + 1/41 + 1/43 + 1/59 + 1/61 + ...$, les dénominateurs sont nombres premiers jumeaux est convergente où finie. *Bull. Sci. Math. 43*, 124-128, 1919.

[3] Chen, J. On the representation of a larger even integer as the sum of a prime and the product of at most two primes. *Sci. Sinica 16*, 157-176, 1973.

[4] Goldston, A., Motohashi Y., Pintz, J. e Yildirim, Y. Small Gaps between Primes Exist. *http://front.math.ucdavis.edu/math.NT/0505300.*

[5] Green, B. e Tao, T. The primes contain arbitrarily long arithmetic progressions. To appear, Annals of Math. http://front.math.ucdavis.edu/math.NT/0404188 (2004).

[6] Goldston, A. e Yildirim, Y. Small Gaps Between Primes I. *http://front.math.ucdavis.edu/math.NT/0504336.*

[7] Roth, K. On certain sets of integers. *J. London Math. Soc. 28* 245-252. 1953.

[8] Schnirelman, L. G. *Uspekhi Math. Nauk 6*, 3-8, 1939.

[9] Szemerédi, E. On sets of integers containing no k elements in arithmetic progression. *Acta Arith. 27* 299-345, 1975.

[10] T. Tao. Arithmetic Progressions and the Primes - El Escorial Lectures. *Lecture Notes*, 2005.

[11] Tao, T. The ergodic and combinatorial approaches to Szemerédi's theorem. *Intended for proceedings of the Montreal workshop on additive combinatorics and number theory.* http://front.math.ucdavis.edu/math.CO/0604456.

[12] Tao, T. The Gaussian primes contain arbitrarily shaped constellations. *J. d´Analyse Mathematique 99*, 109-176 2006.

[13] Tao, T. e Ziegler, T. The primes contain arbitrarily long polynomial progressions. To appear, Acta Math. http://arxiv.org/abs/math.NT/0610050.

[14] Van der Corput, J. Über Summen von Primzahlen und Primzahlquadraten. *Math. Ann. 116*, 1-50, 1939.

[15] Vinogradov, I. Representation of an odd prime as a sum of three primes. *Comptes Rendus (Doklady) de l´Académie des Sciences de PURSS 15* 291-294. 1937.