

Escola Politécnica-USP  
LARC



## Seminário NTT

Resumo

# Contents

<b>1</b>	<b>A Dualidade entre Tempo e Frequência</b>	<b>2</b>
<b>2</b>	<b>Transformada de Fourier Contínua (CTFT)</b>	<b>2</b>
<b>3</b>	<b>Transformada Discreta de Fourier (DFT)</b>	<b>2</b>
3.1	Definição da IDFT . . . . .	3
<b>4</b>	<b>A Multiplicação de Polinômios e a Complexidade Computacional</b>	<b>4</b>
<b>5</b>	<b>A Fast Fourier Transform</b>	<b>5</b>
<b>6</b>	<b>Problemas da FFT</b>	<b>7</b>
<b>7</b>	<b>Number Theoretic Transform (NTT)</b>	<b>7</b>
7.1	Fundamentos . . . . .	7
<b>8</b>	<b>O Isomorfismo via Teorema Chinês dos Restos (CRT)</b>	<b>11</b>
<b>9</b>	<b>Transformada Numérica Inversa (INTT)</b>	<b>12</b>
9.1	A NTT como uma FFT: Decomposição Radix-2 via CRT . . . . .	13
9.2	A INTT via Gentleman-Sande (GS): Reconstrução via CRT . . . . .	15
<b>10</b>	<b>Apêndice</b>	<b>18</b>
10.1	Exemplo de convolução circular . . . . .	18
10.2	Determinição unívoca do polinomio . . . . .	20

# 1 A Dualidade entre Tempo e Frequência

A Transformada de Fourier é uma operação matemática que mapeia uma função do domínio do tempo (ou espaço) para o seu domínio dual: a frequência. Essa transição é extremamente útil, pois propriedades que são complexas de analisar no tempo tornam-se claras no espectro de frequências.

Esta ferramenta é um pilar fundamental em diversas áreas do conhecimento:

- **Matemática Pura:** Essencial na Teoria Analítica dos Números e no estudo de Equações Diferenciais Parciais (EDPs).
- **Física Moderna:** É a base matemática do **Princípio da Incerteza de Heisenberg** na Mecânica Quântica, onde a posição e o momento de uma partícula formam um par de variáveis conjugadas de Fourier.
- **Engenharia:** Processamento de sinais, compressão de dados (MP3, JPEG) e telecomunicações.

## 2 Transformada de Fourier Contínua (CTFT)

Para uma função contínua  $g(t)$ , a transformada é definida pela integral:

$$\mathcal{F}(f) = \int_{-\infty}^{\infty} g(t)e^{-2\pi ift} dt$$

Ela pode ser entendida como um produto interno (uma projeção) do sinal com todos as frequências da reta real  $\langle g, e^{-2\pi ift} \rangle$ , que, devido a ortogonalidade das frequências diferentes e que funções bem comportadas podem ser decompostas em séries de autofunções  $e^{-ift}$ , consegue extrair exatamente as frequências do sinal. Apesar de sua elegância teórica, a CTFT apresenta desafios para a aplicação prática em sistemas digitais:

1. **Natureza Analítica:** A resolução de integrais impróprias exige uma manipulação simbólica que é difícil de implementar em computadores comuns.
2. **Limite Infinito:** A definição pressupõe que conhecemos o sinal de  $-\infty$  a  $+\infty$ , o que é impossível em cenários reais.
3. **Amostragem Finita:** Na prática, os sinais são capturados de forma discreta (amostras) e por um tempo limitado, o que torna a integral contínua inaplicável.

## 3 Transformada Discreta de Fourier (DFT)

Para viabilizar o processamento em computadores, utilizamos a **DFT**. Ela opera sobre uma sequência finita de  $N$  amostras, mapeando dados discretos no tempo para dados discretos na frequência, devido ao teorema de amostragem de Nyquist-Shannon [colocar uma referência], um sinal de frequência  $f$  precisa de  $N \geq f/2$  para ser unicamente determinado.

A DFT é definida da seguinte forma:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-i \frac{2\pi}{N} kn}$$

Para  $k = 0, 1, \dots, N-1$ .

Diferente da versão contínua, a DFT lida com somatórios e vetores numéricos, permitindo que a teoria de Fourier seja aplicada em qualquer dispositivo digital. é possível provar que a DFT é um transformacao linear, logo, pode ser representada matricialmente.

Seja  $\zeta_N = e^{-i \frac{2\pi}{N}}$ . A representação matricial da DFT para  $n = 0, 1, \dots, N-1$  é:

$$\begin{bmatrix} X[0] \\ X[1] \\ \vdots \\ X[N-1] \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \zeta_N^1 & \zeta_N^2 & \dots & \zeta_N^{N-1} \\ 1 & \zeta_N^2 & \zeta_N^4 & \dots & \zeta_N^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \zeta_N^{N-1} & \zeta_N^{2(N-1)} & \dots & \zeta_N^{(N-1)(N-1)} \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix}$$

### 3.1 Definição da IDFT

A reconstrução do sinal original no domínio do tempo a partir de suas amostras de frequência é realizada pela IDFT:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] \zeta_N^{-nk}, \quad n = 0, 1, \dots, N-1$$

Onde  $\zeta_N^{-nk} = e^{i \frac{2\pi}{N} nk}$ . Matricialmente, a IDFT é dada por:

$$\begin{bmatrix} x[0] \\ x[1] \\ x[2] \\ \vdots \\ x[N-1] \end{bmatrix} = \frac{1}{N} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \zeta_N^{-1} & \zeta_N^{-2} & \dots & \zeta_N^{-(N-1)} \\ 1 & \zeta_N^{-2} & \zeta_N^{-4} & \dots & \zeta_N^{-2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \zeta_N^{-(N-1)} & \zeta_N^{-2(N-1)} & \dots & \zeta_N^{-(N-1)(N-1)} \end{bmatrix} \begin{bmatrix} X[0] \\ X[1] \\ X[2] \\ \vdots \\ X[N-1] \end{bmatrix}$$

Percebe-se que nesse caso discreto, a DFT e IDFT atuam como uma matriz mudanca de base, saindo da base do tempo e indo para base das raízes unitarias

A DFT pde ser vista de outra óptica, como avaliação de polinomios. Seja

$$a(x) = \sum_{n=0}^{N-1} a_n x^n$$

e tome  $\zeta_N = e^{-\frac{2\pi i}{N}}$  seja uma  $N$ -ésima raiz primitiva da unidade. Defina

$$A_k := a(\zeta_N^k).$$

Então

$$A_k = \sum_{n=0}^{N-1} a_n (\zeta_N^k)^n = \sum_{n=0}^{N-1} a_n \zeta_N^{kn},$$

o que é exatamente a formulação da DFT, para  $a_n \equiv x[n]$  e  $A_n \equiv X[n]$ , opostamente, a IDFT é interpretada como interpolação de polinômios:

Dados os valores  $\{A_k\}_{k=0}^{N-1}$ , ela reconstrói os coeficientes  $\{a_n\}_{n=0}^{N-1}$  do único polinômio de grau  $N - 1$  que satisfaz  $a(\zeta_N^k) = A_k$  para todo  $k$ . Explicitamente,

$$a_n = \frac{1}{N} \sum_{k=0}^{N-1} A_k \zeta_N^{-kn}.$$

Assim, DFT é *avaliar* em raízes da unidade e IDFT é *interpolar* (recuperar os coeficientes) a partir dessas avaliações. (Essa perspectiva baseia-se na determinição unívoca do polinômio de grau  $N - 1$  por  $N$  pontos veja a seção 10.2 para demonstração desse fato)

## 4 A Multiplicação de Polinômios e a Complexidade Computacional

Um problema simplificado pela mudança de domínio é a multiplicação de polinômios. Tome os polinômios  $f(x)$  e  $g(x)$  de grau  $N - 1$ :

$$f(x) = \sum_{i=0}^{N-1} a_i x^i, \quad g(x) = \sum_{j=0}^{N-1} b_j x^j$$

Na abordagem clássica, o produto  $h(x) = f(x) \cdot g(x)$  é obtido distribuindo-se cada termo de  $f$  sobre todos os termos de  $g$ . Este processo resulta em um novo polinômio de grau  $2N - 2$ :

$$h(x) = \sum_{k=0}^{2N-2} c_k x^k$$

onde  $c_k = \sum_{i+j=k} a_i b_j$ .

Nesta metodologia, o cálculo de cada coeficiente  $c_k$  exige múltiplas operações de produto e soma, resultando em uma complexidade assintótica  $O(n^2)$ . Para polinômios com grandes volumes de coeficientes, este custo computacional torna o método inviável.

A conexão entre a multiplicação de polinômios e a análise de Fourier vem do fato de que, se  $h(x) = f(x)g(x)$ , então os coeficientes  $c_k$  de  $h$  são dados pela **convolução linear** dos coeficientes de  $f$  e  $g$ :

$$c_k = \sum_{i+j=k} a_i b_j.$$

Neste trabalho, devido ao foco na NTT e ao anel quociente  $\mathbb{Z}_p[x]/(x^N + 1)$ , trabalhamos com a **convolução circular** (de comprimento  $n$ ), definida por

$$c_k = \sum_{i=0}^{N-1} a_i b_{(k-i) \bmod n}, \quad k = 0, \dots, N - 1.$$

**Teorema da Convolução:** a transformada de uma convolução no domínio do tempo (ou espaço) é o produto ponto a ponto (Hadamard) das transformadas no domínio da frequência:

$$\mathcal{F}(f * g) = \mathcal{F}(f) \odot \mathcal{F}(g).$$

Assim, o cálculo custoso da convolução é convertido em um produto ponto a ponto, pois a base de Fourier diagonaliza o operador de convolução (circulante). Com a transformada direta ingênua o custo ainda é  $O(n^2)$ , não há ganho da multiplicação clássica de polinômios.

Uma exemplo para o convencimento do leitor foi disposto no apêndice 10.1

## 5 A Fast Fourier Transform

A FFT (Fast Fourier Transform) é uma maneira de otimizar o cálculo da DFT.

O algoritmo da FFT foi redescoberta por Cooley e Tukey em 1965, uma vez que Gauss já tinha utilizado um algoritmo semelhante para calcular as órbitas de asteroides em 1805.

O algoritmo se baseia em **dividir para conquistar**.

Relembrando

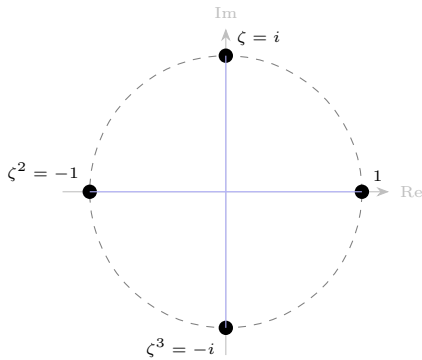
As raízes unitárias possuem propriedades cíclicas e certas simetrias que permitem a economia nos cálculos, vejamos um exemplo.

$$\zeta_4^1 = e^{-i\frac{2\pi}{4}} = e^{-i90^\circ} = -i$$

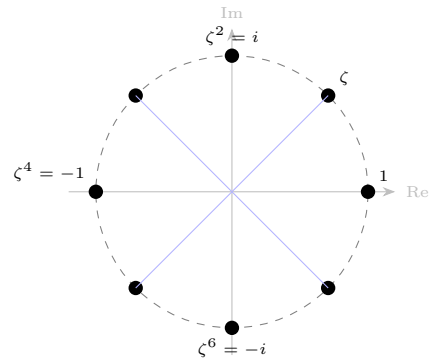
$$\zeta^1 = -i \quad \zeta^2 = -1$$

$$\zeta^3 = i \quad \zeta^4 = 1$$

por isso, percebe-se que, a cada 2 "deslocamentos", o valor se torna o oposto, como ilustrado na figura:



(a) 4-ésimas raízes da unidade



(b) 8-ésimas raízes da unidade

Figure 1: Comparação entre as raízes da unidade no plano complexo.

De forma mais geral:

$$\zeta_N = e^{\frac{-2\pi i}{N}}, \text{ uma raiz } N\text{-ésima primitiva da unidade. Então para todo inteiro } a,$$

$$\zeta_N^{a+\frac{N}{2}} = -\zeta_N^a.$$

Alem de que, pela periodicidade  $\zeta_N^{a+N} = \zeta_N^a$ .

Para esse caso a DFT é representada desse modo:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{bmatrix}$$

Voltando a FFT, o algoritmo decompõe uma DFT de tamanho  $N$  em duas sub-transformadas de tamanho  $N/2$ , separando os índices pares e ímpares da sequência original:

$$X[k] = \sum_{m=0}^{\frac{N}{2}-1} x[2m] \zeta_N^{2mk} + \sum_{m=0}^{\frac{N}{2}-1} x[2m+1] \zeta_N^{(2m+1)k}$$

$$\text{Usando } \zeta_N^2 = \zeta_{N/2} : \quad \zeta_N^{2mk} = (\zeta_N^2)^{mk} = \zeta_{N/2}^{mk}$$

$$\begin{aligned} &= \sum_{m=0}^{\frac{N}{2}-1} x[2m] \zeta_{N/2}^{mk} + \zeta_N^k \sum_{m=0}^{\frac{N}{2}-1} x[2m+1] \zeta_{N/2}^{mk} \\ &= E[k] + \zeta_N^k O[k], \quad k = 0, \dots, \frac{N}{2} - 1. \end{aligned}$$

Esta estrutura permite calcular dois valores de saída ( $X[k]$  e  $X[k + N/2]$ ) utilizando os mesmos resultados intermediários, através da denominada **operação borboleta** (*butterfly operation*):

1.  $X[k] = E[k] + \zeta_N^k O[k]$
2.  $X[k + N/2] = E[k] - \zeta_N^k O[k]$

como pode ser visto na imagem

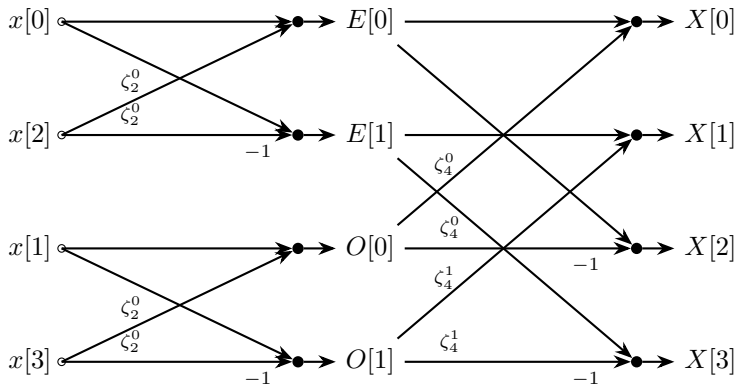


Figure 2: butterfly radix-2

o a implementacao em sage esta no codigo1

Listing 1: Implementação do algoritmo FFT em SageMath

```

1 def fft(a, omega):
2     n = len(a)
3     if n == 1:
4         return a[:]
5
6
7     a_par = fft(a[0::2], omega^2)
8     a_impar = fft(a[1::2], omega^2)
9
10    A = [0] * n
11    w = 1
12    half = n // 2
13    for k in range(half):
14        t = w * a_impar[k]
15        A[k] = a_par[k] + t
16        A[k + half] = a_par[k] - t
17        w *= omega
18    return A

```

Desse modo, reduzimos a complexidade da transformada de  $O(n^2)$  para  $O(n \cdot \log n)$ . Por causa disso, podemos utilizar a FFT, junto com o **teorema da convolucao**, para multiplicar polinomios em  $O(n \cdot \log n)$

## 6 Problemas da FFT

Uns dos problemas da FFT é que ela trabalha com ponto flutuante, o que, para computadores, é um grande problemas que pode causar erro de arredondamentos e, assim causar uma falha nos esquemas criptograficos. Alem disso, na convolução linear, o polinomio dobra de tamanho a cada concolucao o que rapidamente torna-se um problema tanto computacional quanto de armazenamento .

Solucao: utilizar um transformada que utiliza apenas numeros exatos

## 7 Number Theoretic Transform (NTT)

### 7.1 Fundamentos

As propriedades que usamos na FFT — em especial a existência de uma raiz  $N$ -ésima da unidade  $\zeta_N$  e o fato de que suas potências percorrem uniformemente o círculo — têm um análogo perfeito em teoria dos números, dentro de corpos (ou anéis) finitos. Isso não é coincidência: a FFT nada mais é do que a transformada de Fourier no grupo cíclico  $\mathbb{Z}/N\mathbb{Z}$ , e a mesma construção existe em outros contextos algébricos.

Mais formalmente, se  $\zeta_N$  é uma raiz  $N$ -ésima primitiva da unidade, então o conjunto de todas as  $N$ -ésimas raízes

$$\mu_N = \{1, \zeta_N, \zeta_N^2, \dots, \zeta_N^{N-1}\}$$



forma um grupo multiplicativo cíclico de ordem  $N$ . Existe um isomorfismo natural de grupos

$$\varphi : \mathbb{Z}/N\mathbb{Z} \rightarrow \mu_N, \quad \varphi([k]) = \zeta_N^k,$$

onde o lado esquerdo usa a soma módulo  $N$  e o lado direito usa multiplicação:

$$\varphi([k + \ell]) = \zeta_N^{k+\ell} = \zeta_N^k \zeta_N^\ell = \varphi([k]) \varphi([\ell]).$$

## Raiz Primitiva e Estrutura Negacíclica

Diferente da DFT complexa, onde raízes da unidade sempre existem para qualquer  $N$ , a NTT exige que o corpo finito  $\mathbb{Z}_p$ , isto é, os inteiros  $(\text{mod } p)$ , sendo  $p$  um primo, suporte a ordem da transformada.

Para a NTT Negacíclica, utilizada para realizar a multiplicação polinomial módulo  $x^N + 1$ , precisamos de uma raiz primitiva  $2N$ -ésima da unidade em  $\mathbb{Z}_p$ , que denotaremos por  $\psi$ . Isso implica que:

1.  $\psi^{2N} \equiv 1 \pmod{p}$ ;
2.  $\psi^N \equiv -1 \pmod{p}$ .

Para garantir a existência desse elemento, a ordem  $2N$  deve dividir a ordem do grupo multiplicativo do corpo. Portanto, o primo  $p$  deve satisfazer:

$$2N \mid (p - 1) \quad \text{i.e. } 2N \text{ divide } (p-1)$$

Ou seja,  $p \equiv 1 \pmod{2N}$ .

Como  $2N \mid (p - 1)$ , em particular temos  $p \nmid N$  (i.e.  $\gcd(p, N) = 1$ ); logo  $N$  possui inverso em  $\mathbb{Z}_p$ .

Agora, trataremos da estrutura algébrica. A NTT Negacíclica é definida no anel quociente

$$R = \frac{\mathbb{Z}_p[x]}{(x^N + 1)}$$

(a imagem 3 ilustra a sequencia das duas operacoes visualmente.)

Ela pode ser vista como a avaliação do polinômio nas raízes da equação  $x^N + 1 = 0$ , que correspondem às potências ímpares de  $\psi$ . A transformada é definida por:

$$X[k] = \sum_{n=0}^{N-1} x[n] \psi^{(2k+1)n} \pmod{p}$$

Na forma matricial:

$$\mathbf{X} = \mathbf{W}_N \mathbf{x}, \quad \mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix}$$

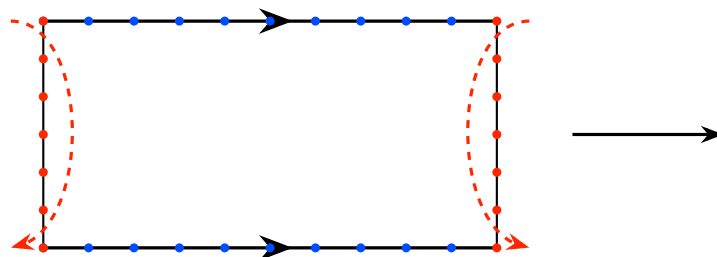
Neste caso, a matriz de transformação  $\mathbf{W}_N$  difere da versão cíclica padrão, pois seus coeficientes seguem a estrutura das raízes negacíclicas:

$$\mathbf{W}_N = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \psi^3 & \psi^6 & \dots & \psi^{3(N-1)} \\ 1 & \psi^5 & \psi^{10} & \dots & \psi^{5(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \psi^{(2k+1)} & \psi^{(2k+1)2} & \dots & \psi^{(2k+1)(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \psi^{(2N-1)} & \psi^{(2N-1)2} & \dots & \psi^{(2N-1)(N-1)} \end{bmatrix}$$

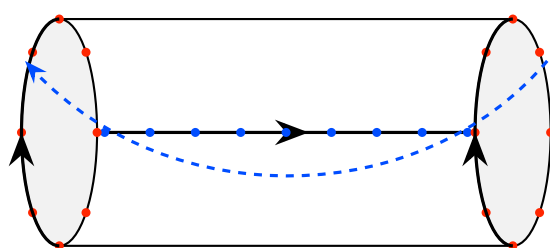
De forma geral, o termo na linha  $k$  e coluna  $n$  da matriz é dado por:

$$(\mathbf{W}_N)_{k,n} = \psi^{(2k+1)n} \pmod{p}, \quad 0 \leq k, n \leq N-1$$

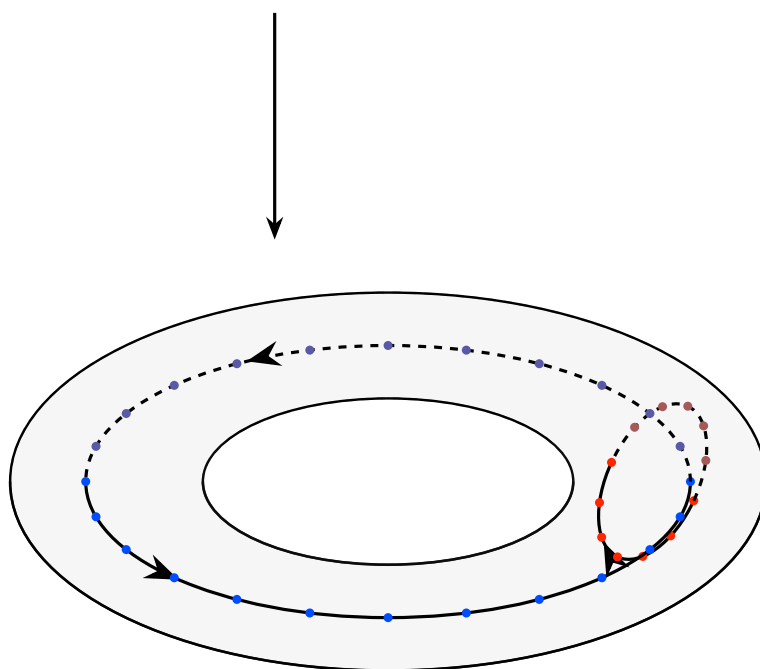
1. Colar lados horizontais (Passo  $\mathbb{Z}_p$ )



Obtém-se um cilindro



2. Colar lados verticais (Passo  $/(x^N + 1)$ )



3. Obtém-se um Toro Discreto

• Direção  $/(x^N + 1)$  • Direção  $\mathbb{Z}_p$

Figure 3: representação das transformacoes

## 8 O Isomorfismo via Teorema Chinês dos Restos (CRT)

A fundamentação algébrica da NTT reside na estrutura do anel quociente  $\mathbb{Z}_p[x]/\langle x^N + 1 \rangle$ . Dado que  $x^N + 1$  fatora-se em  $N$  binômios lineares distintos  $(x - \psi^{2k+1})$ , o CRT estabelece uma equivalência entre o anel original e o produto de anéis menores.

Para compreender a natureza desses anéis menores, invocamos o conceito de *homomorfismo*. Considere o anel quociente genérico  $\mathbb{Z}_p[x]/\langle x - \alpha \rangle$ . Pela Divisão Euclidiana, qualquer polinômio  $a(x) \in \mathbb{Z}_p[x]$  pode ser escrito unicamente como:

$$a(x) = q(x) \cdot (x - \alpha) + r$$

onde o grau de  $r$  deve ser estritamente menor que o grau do divisor  $(x - \alpha)$ . Como o divisor tem grau 1, o resto  $r$  é necessariamente um escalar constante ( $r \in \mathbb{Z}_p$ ).

No anel quociente, impomos que  $x - \alpha \equiv 0$ , ou seja,  $x \equiv \alpha$ . Ao aplicarmos essa equivalência (ou avaliarmos a equação em  $x = \alpha$ ), o termo que contém o divisor se anula:

$$a(\alpha) = q(\alpha) \cdot \underbrace{(\alpha - \alpha)}_0 + r \implies a(\alpha) = r$$

Portanto, a classe de equivalência de  $a(x)$  módulo  $(x - \alpha)$  é univocamente representada pelo escalar  $a(\alpha)$ . Isso estabelece o isomorfismo local:

$$\frac{\mathbb{Z}_p[x]}{\langle x - \alpha \rangle} \cong \mathbb{Z}_p \quad \text{via o mapa} \quad [a(x)] \mapsto a(\alpha)$$

Com essa intuição firmada, podemos expandir explicitamente o isomorfismo global  $\Phi$  garantido pelo CRT. A transformação mapeia o polinômio  $a(x)$  diretamente para o vetor de suas avaliações nas raízes de  $x^N + 1$ :

$$\begin{aligned} \Phi : \frac{\mathbb{Z}_p[x]}{\langle x^N + 1 \rangle} &\xrightarrow{\cong} \bigotimes_{k=0}^{N-1} \frac{\mathbb{Z}_p[x]}{\langle x - \psi^{2k+1} \rangle} \\ &\cong \underbrace{\mathbb{Z}_p \times \mathbb{Z}_p \times \cdots \times \mathbb{Z}_p}_{N \text{ vezes}} \end{aligned}$$

A ação de  $\Phi$  sobre um elemento  $a(x)$  é dada explicitamente por:

$$\Phi(a(x)) = (a(\psi^1), a(\psi^3), \dots, a(\psi^{2N-1})) \in \mathbb{Z}_p^N$$

Neste domínio transformado, a operação de multiplicação de dois polinômios  $a(x)$  e  $b(x)$  ocorre componente a componente. Se  $C = \Phi(a(x) \cdot b(x))$ , então a  $k$ -ésima componente do vetor resultante é:

$$C_k = (a \cdot b)(\psi^{2k+1}) = a(\psi^{2k+1}) \cdot b(\psi^{2k+1})$$

Isso demonstra formalmente por que a convolução cíclica (ou negacíclica) no domínio do tempo se traduz em um produto de Hadamard (ponto a ponto) no domínio da frequência.

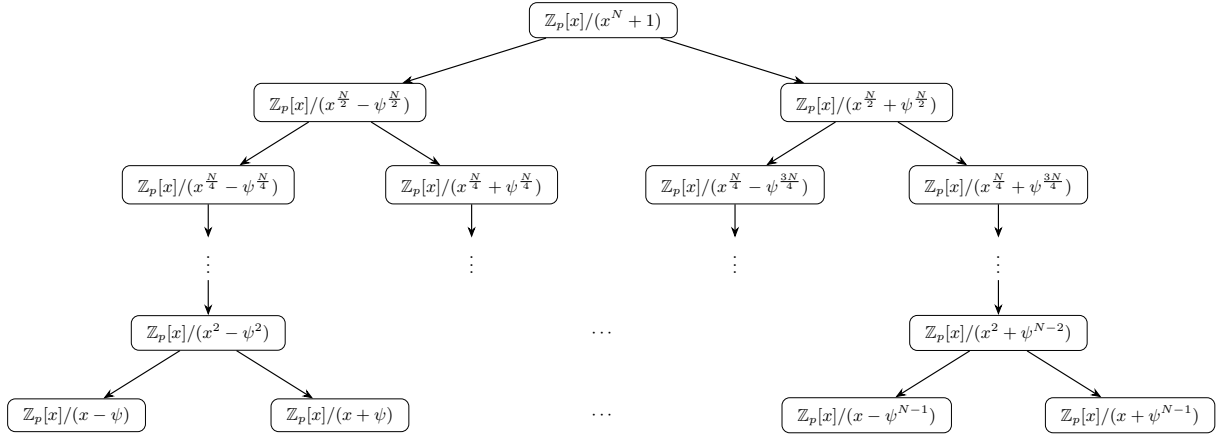


Figure 4: CRT da NTT Negacíclica

## 9 Transformada Numérica Inversa (INTT)

Como o Teorema Chinês dos Restos garante que a aplicação da NTT é um isomorfismo bijetor entre o anel de polinômios  $\mathbb{Z}_p[x]/(x^N + 1)$  e o domínio da frequência, existe uma transformação inversa única capaz de recuperar os coeficientes originais.

Denotamos a inversa multiplicativa de  $N$  no corpo  $\mathbb{Z}_p$  por  $N^{-1}$ , tal que  $N \cdot N^{-1} \equiv 1 \pmod{p}$ . A \*\*NTT Negacíclica Inversa (INTT)\*\* é definida formalmente por:

$$x[n] = N^{-1} \sum_{k=0}^{N-1} X[k] \psi^{-(2k+1)n} \pmod{p}$$

Note que o termo  $\psi^{-(2k+1)n}$  refere-se à potência do inverso multiplicativo da raiz, ou seja,  $\psi^{-1} \equiv \psi^{2N-1} \pmod{p}$ .

### Representação Matricial

Na forma matricial, a operação de inversão corresponde à resolução do sistema linear  $\mathbf{X} = \mathbf{W}_N \mathbf{x}$ . A solução é dada por:

$$\mathbf{x} = \mathbf{W}_N^{-1} \mathbf{X}$$

Onde a matriz inversa  $\mathbf{W}_N^{-1}$  é definida como a conjugada transposta da matriz direta, escalada pelo fator  $N^{-1}$ :

$$\mathbf{W}_N^{-1} = N^{-1} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \psi^{-3} & \psi^{-6} & \dots & \psi^{-(2N-1)} \\ 1 & \psi^{-6} & \psi^{-10} & \dots & \psi^{-2(2N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \psi^{-(2k+1)} & \psi^{-2(2k+1)} & \dots & \psi^{-(N-1)(2k+1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \psi^{-(2N-1)} & \psi^{-2(2N-1)} & \dots & \psi^{-(N-1)(2N-1)} \end{bmatrix}$$

De modo geral, o termo na linha  $n$  e coluna  $k$  da matriz inversa é:

$$(\mathbf{W}_N^{-1})_{n,k} = N^{-1} \psi^{-(2k+1)n} \pmod{p}$$

## Ortogonalidade

A existência da inversa e a validade da fórmula acima decorrem da propriedade de ortogonalidade das raízes negacíclicas no corpo finito  $\mathbb{Z}_p$ . Para quaisquer  $n, m \in \{0, \dots, N-1\}$ , temos:

$$\sum_{k=0}^{N-1} \psi^{(2k+1)(n-m)} = \begin{cases} N & \text{se } n = m \\ 0 & \text{se } n \neq m \end{cases}$$

Isso garante que o produto das matrizes resulta na identidade:

$$\mathbf{W}_N^{-1} \mathbf{W}_N = \mathbf{I}$$

### 9.1 A NTT como uma FFT: Decomposição Radix-2 via CRT

A eficiência da NTT reside na estratégia “dividir para conquistar”. Em vez de avaliar um polinômio de grau  $N-1$  de uma só vez, utilizamos a estrutura algébrica do anel para decompor o problema em dois problemas independentes de tamanho  $N/2$ .

#### 1. A Fatoração do Módulo

O ponto de partida é a decomposição do polinômio  $X^N + 1$ . Como  $\psi$  é uma raiz primitiva  $2N$ -ésima, temos que  $\psi^N \equiv -1$ . Isso implica que o termo  $\psi^{N/2}$  atua como uma “raiz quadrada de -1” no corpo finito.

Podemos então ver  $X^N + 1$  como uma diferença de quadrados:

$$X^N + 1 = X^N - (\psi^{N/2})^2 = (X^{N/2} - \psi^{N/2}) \cdot (X^{N/2} + \psi^{N/2})$$

O Teorema Chinês dos Restos (CRT) garante que resolver o problema no anel original é equivalente a resolvê-lo simultaneamente nestes dois sub-anéis menores.

## 2. O Colapso da Variável

Considere o polinômio de entrada  $A(X)$  dividido em sua metade inferior ( $A_{lo}$ ) e superior ( $A_{hi}$ ):

$$A(X) = \sum_{n=0}^{N/2-1} x[n]X^n + X^{N/2} \sum_{n=0}^{N/2-1} x[n + N/2]X^n$$

A intuição chave da NTT é que, ao projetarmos  $A(X)$  nos sub-anéis definidos pelos fatores acima, a potência  $X^{N/2}$  deixa de ser uma variável e colapsa em uma constante:

- **No ramo esquerdo** (módulo  $X^{N/2} - \psi^{N/2}$ ): Temos a equivalência  $X^{N/2} \equiv \psi^{N/2}$ .
- **No ramo direito** (módulo  $X^{N/2} + \psi^{N/2}$ ): Temos a equivalência  $X^{N/2} \equiv -\psi^{N/2}$ .

Isso simplifica drasticamente o cálculo, pois reduz a multiplicação polinomial a somas ponderadas.

## 3. A Operação Borboleta (Butterfly)

Substituindo as equivalências acima na expressão de  $A(X)$ , obtemos as fórmulas para os novos polinômios  $A_L(X)$  e  $A_R(X)$  de grau  $N/2 - 1$ . Em termos dos coeficientes, para cada  $0 \leq i < N/2$ :

$$\begin{cases} x_L[i] = x[i] + \psi^{N/2} \cdot x[i + N/2] \pmod{p} \\ x_R[i] = x[i] - \psi^{N/2} \cdot x[i + N/2] \pmod{p} \end{cases}$$

Esta é a clássica operação “borboleta”. Observe a simetria: os mesmos valores de entrada ( $x[i]$  e  $x[i + N/2]$ ) são usados para calcular ambas as saídas, diferenciando-se apenas pelo sinal da operação.

O processo é então aplicado recursivamente:  $x_L$  torna-se a entrada para uma NTT de tamanho  $N/2$ , e  $x_R$  para outra, até que o tamanho do problema seja trivial (1).

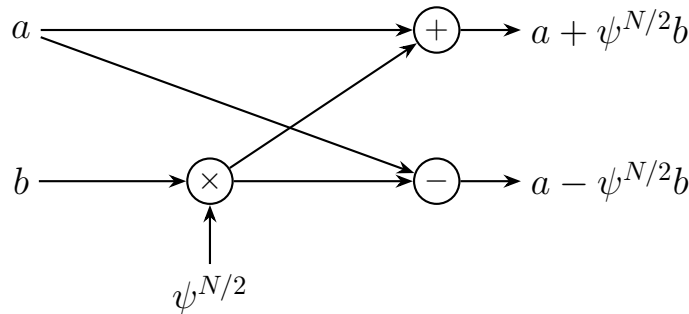


Figure 5: Butterfly NTT

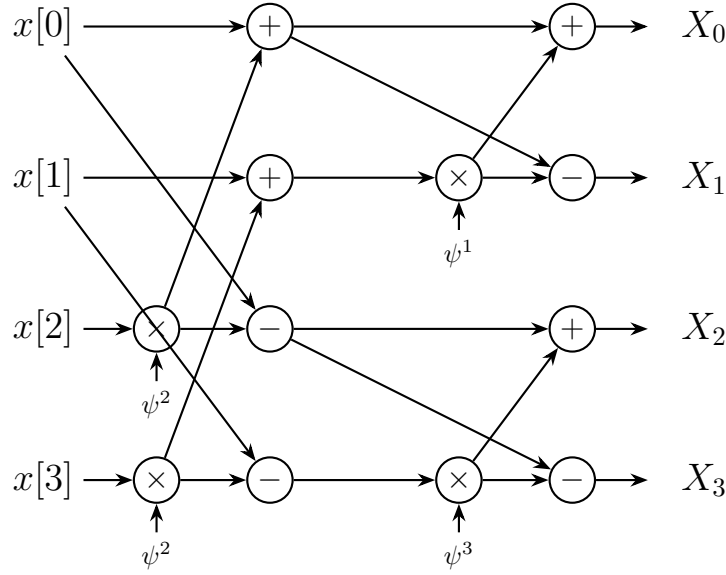


Figure 6: Butterfly NTT

## 9.2 A INTT via Gentleman-Sande (GS): Reconstrução via CRT

Enquanto a transformação direta (Forward NTT) utiliza a estratégia de decomposição para reduzir o grau do polinômio, a transformação inversa (INTT) realiza o caminho oposto: ela combina pares de polinômios menores para reconstruir o polinômio original modulo  $X^N + 1$ . O algoritmo de Gentleman-Sande implementa essa reconstrução de forma eficiente.

### 1. O Problema da Interpolação

Na etapa inversa, começamos com dois polinômios "metade" recuperados dos sub-anéis:

- $A_L(X)$ : O resultado associado ao ramo  $X^{N/2} - \psi^{N/2}$ .
- $A_R(X)$ : O resultado associado ao ramo  $X^{N/2} + \psi^{N/2}$ .

Nosso objetivo é encontrar o polinômio único  $A(X) = A_{lo}(X) + X^{N/2}A_{hi}(X)$  que satisfaz:

$$\begin{cases} A(X) \equiv A_L(X) \pmod{X^{N/2} - \psi^{N/2}} \\ A(X) \equiv A_R(X) \pmod{X^{N/2} + \psi^{N/2}} \end{cases}$$

### 2. A Lógica da Recuperação

Lembrando que no sub-anel esquerdo  $X^{N/2} \equiv \psi^{N/2}$  e no direito  $X^{N/2} \equiv -\psi^{N/2}$ , podemos escrever o sistema linear para os componentes  $A_{lo}$  e  $A_{hi}$ :

$$\begin{aligned} A_L &= A_{lo} + \psi^{N/2} A_{hi} \\ A_R &= A_{lo} - \psi^{N/2} A_{hi} \end{aligned}$$

Para isolar  $A_{lo}$  e  $A_{hi}$ , basta somar e subtrair as equações:



- **Soma:**  $A_L + A_R = 2A_{lo} \implies A_{lo} = \frac{1}{2}(A_L + A_R)$
- **Subtração:**  $A_L - A_R = 2\psi^{N/2}A_{hi} \implies A_{hi} = \frac{1}{2}\psi^{-N/2}(A_L - A_R)$

Esta estrutura algébrica revela a diferença fundamental entre a borboleta da ida (Cooley-Tukey) e a da volta (Gentleman-Sande).

### 3. A Borboleta GS (Inverse Butterfly)

Diferente da NTT direta, onde a multiplicação pelo fator de rotação ocorre *antes* da borboleta (ou no ramo inferior da entrada), na INTT via Gentleman-Sande a multiplicação ocorre *após* a subtração.

Para coeficientes de entrada  $y_L$  e  $y_R$  correspondentes à mesma posição nos vetores dos sub-blocos, a operação de fusão é:

$$\begin{aligned} x[i] &= \frac{1}{2}(y_L[i] + y_R[i]) \pmod{p} \\ x[i + N/2] &= \frac{1}{2}(y_L[i] - y_R[i]) \cdot \psi^{-N/2} \pmod{p} \end{aligned}$$

*Nota: Em implementações eficientes, o fator escalar  $1/2$  (inverso modular de 2) é frequentemente acumulado e aplicado apenas uma vez no final como  $1/N$ , para evitar multiplicações excessivas a cada estágio.*

### 4. Fluxo de Dados

O algoritmo Gentleman-Sande é a "transposta" do Cooley-Tukey. Se a NTT direta recebe os dados em ordem natural e os devolve em ordem *bit-reversed* (embaralhada), a INTT via GS aceita os dados em ordem *bit-reversed* e os reconstrói para a ordem natural. Isso elimina a necessidade de reordenamentos custosos de memória entre a transformação direta e a inversa durante operações de convolução.

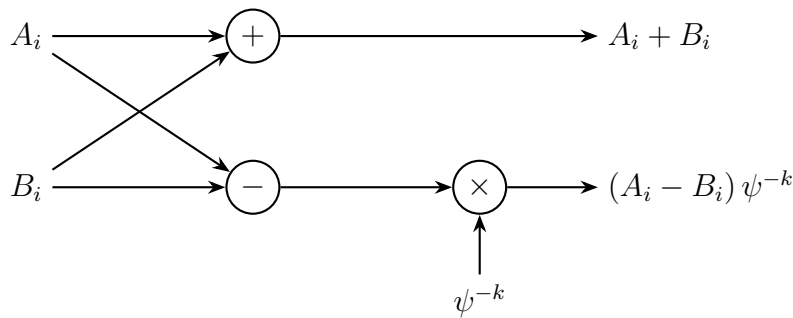


Figure 7: Butterfly INTT

A implementa da NTT negacíclica é a seguir 2

Listing 2: código em SageMath

```

1 def NTT_Negacyclic(A, R, psi):
2
3     n = len(A)
4     if n == 1:
5         return [A[0]]
6
7     A_even = A[0::2]
8     A_odd  = A[1::2]
9
10    psi_sq = psi^2
11    Y_even = NTT_Negacyclic(A_even, R, psi_sq)
12    Y_odd  = NTT_Negacyclic(A_odd, R, psi_sq)
13
14    Y = [R(0)] * n
15
16    w = psi
17
18    half = n // 2
19    for k in range(half):
20        t = w * Y_odd[k]
21
22        Y[k]          = Y_even[k] + t
23        Y[k + half] = Y_even[k] - t
24
25        w *= psi_sq
26
27    return Y
28
29 def INTT_Negacyclic(A, R, psi):
30     n = len(A)
31     # A inversa usa psi^-1
32     psi_inv = psi^(-1)
33
34     Y = NTT_Negacyclic(A, R, psi_inv)
35
36     n_inv = R(n)^(-1)
37     return [y * n_inv for y in Y]

```

A comporação computacional dessa ferramentas pode ser vista na tabela a seguir, onde calculou-se os números de Fibonacci.

Algorithm	Fibonacci index
Algoritmo Naive	44
Algoritmo Linear	566'053
Algoritmo FFT	3'145'816
Algoritmo NTT	24'178'839
Algoritmo GMP	238'961'323

## 10 Apêndice

### 10.1 Exemplo de convolução circular

Sejam dois sinais  $x$  e  $y$  de comprimento  $N = 3$ :

$$x = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad y = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

A frequência fundamental é a raiz da unidade  $\zeta_3 = e^{-i\frac{2\pi}{3}}$ . Usaremos a propriedade fundamental:

$$1 + \zeta_3 + \zeta_3^2 = 0 \implies \zeta_3 + \zeta_3^2 = -1$$

#### Método 1: Convolução no Tempo (Matriz Circulante)

A convolução circular  $z = x \circledast y$  equivale à multiplicação de uma matriz circulante  $C_x$  pelo vetor  $y$ :

$$C_x = \begin{bmatrix} x[0] & x[2] & x[1] \\ x[1] & x[0] & x[2] \\ x[2] & x[1] & x[0] \end{bmatrix} = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}$$

Calculando  $z = C_x y$ :

$$z = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1(1) + 0 + 2(1) \\ 2(1) + 0 + 0 \\ 0 + 0 + 1(1) \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

#### Método 2: Diagonalização (Domínio da Frequência)

Mostra-se que a mudança de base para a base de Fourier diagonaliza a matriz circulante  $C_x$ . Matematicamente, isso significa que  $C_x$  pode ser decomposta como

$$C_x = F^{-1} \Lambda_x F$$

onde  $\Lambda_x$  é uma matriz diagonal contendo os coeficientes da DFT de  $x$ .

Substituindo essa decomposição na equação original da convolução  $z = C_x y$ , podemos rearranjar os termos para utilizar a multiplicação na frequência:

$$z = (F^{-1} \Lambda_x F) y = F^{-1} \Lambda_x (F y) = F^{-1} \Lambda_x Y \quad (1)$$

Dessa forma, o cálculo se resume a obter  $Y$  (DFT de  $y$ ), multiplicar pela matriz diagonal  $\Lambda_x$  e aplicar a inversa ( $F^{-1}$ ).

Para legitimar que os autovalores de  $C_x$  são os coeficientes da DFT de  $x$ , resolvemos:

$$\det(C_x - \lambda I) = 0 \implies \det \begin{bmatrix} 1 - \lambda & 0 & 2 \\ 2 & 1 - \lambda & 0 \\ 0 & 2 & 1 - \lambda \end{bmatrix} = 0$$

Expandindo o determinante:

$$(1 - \lambda)^3 + 8 = 0 \implies (1 - \lambda)^3 = -8$$

As raízes para  $(1 - \lambda)$  são as três raízes cúbicas de  $-8$ :

$$\begin{aligned} 1 - \lambda_0 &= -2 \implies \lambda_0 = 3 \\ 1 - \lambda_1 &= -2\zeta_3 \implies \lambda_1 = 1 + 2\zeta_3 \\ 1 - \lambda_2 &= -2\zeta_3^2 \implies \lambda_2 = 1 + 2\zeta_3^2 \end{aligned}$$

Estes valores coincidem exatamente com a DFT de  $x$ , provando a legitimidade da diagonalização.

## Verificação do Autovetor

Verificamos agora se o autovetor  $v_1$  da base de Fourier (coluna de  $F^{-1}$ ), dado por  $v_1 = [1, \zeta_3^2, \zeta_3]^T$ , satisfaz  $C_x v_1 = \lambda_1 v_1$ .

Lado esquerdo ( $C_x v_1$ ):

$$\begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \zeta_3^2 \\ \zeta_3 \end{bmatrix} = \begin{bmatrix} 1 + 2\zeta_3 \\ 2 + \zeta_3^2 \\ 2\zeta_3^2 + \zeta_3 \end{bmatrix}$$

Lado direito ( $\lambda_1 v_1$ ):

$$(1 + 2\zeta_3) \begin{bmatrix} 1 \\ \zeta_3^2 \\ \zeta_3 \end{bmatrix} = \begin{bmatrix} 1 + 2\zeta_3 \\ \zeta_3^2 + 2\zeta_3^3 \\ \zeta_3 + 2\zeta_3^2 \end{bmatrix} = \begin{bmatrix} 1 + 2\zeta_3 \\ 2 + \zeta_3^2 \\ \zeta_3 + 2\zeta_3^2 \end{bmatrix}$$

A igualdade é satisfeita, pensamento analogo é utilizado para mostrar que o  $v_0$  e  $v_2$  e também são autovetores, confirmando que a base de Fourier é a base natural de  $C_x$ .

## Representação Matricial da Diagonalização

Primeiro, calculamos os vetores transformados  $X = Fx$  e  $Y = Fy$ :

$$X = \begin{bmatrix} 3 \\ 1 + 2\zeta_3 \\ 1 + 2\zeta_3^2 \end{bmatrix}, \quad Y = \begin{bmatrix} 2 \\ 1 + \zeta_3^2 \\ 1 + \zeta_3 \end{bmatrix}$$

Agora, construímos a matriz diagonal  $\Lambda_x = \text{diag}(X)$ . A operação de convolução no domínio da frequência ( $Z = X \cdot Y$ ), visualizada matricialmente como  $Z = \Lambda_x Y$ , torna-se:

$$\mathbf{Z} = \underbrace{\begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 + 2\zeta_3 & 0 \\ 0 & 0 & 1 + 2\zeta_3^2 \end{bmatrix}}_{\text{Matriz Diagonal } (\Lambda_x)} \begin{bmatrix} 2 \\ 1 + \zeta_3^2 \\ 1 + \zeta_3 \end{bmatrix}$$

Executando o produto matricial (que equivale ao produto ponto a ponto):

$$\mathbf{Z} = \begin{bmatrix} 3 \cdot 2 \\ (1 + 2\zeta_3)(1 + \zeta_3^2) \\ (1 + 2\zeta_3^2)(1 + \zeta_3) \end{bmatrix} = \begin{bmatrix} 6 \\ 2 + \zeta_3 \\ 2 + \zeta_3^2 \end{bmatrix}$$

*\*Nota: As simplificações algébricas utilizam  $1 + \zeta_3 + \zeta_3^2 = 0$ .*

## Retorno ao Tempo (IDFT)

Finalmente, aplicamos a matriz inversa de Fourier ( $F^{-1}$ ) para obter  $z$ :

$$z = F^{-1}\mathbf{Z} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & \zeta_3^{-1} & \zeta_3^{-2} \\ 1 & \zeta_3^{-2} & \zeta_3^{-4} \end{bmatrix} \begin{bmatrix} 6 \\ 2 + \zeta_3 \\ 2 + \zeta_3^2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

Isso confirma numericamente que a base de Fourier diagonalizou a operação.

## 10.2 Determinação unívoca do polinômio

**Theorem 1** (Unicidade via matriz de Vandermonde). *Sejam  $x_0, \dots, x_{N-1}$  escalares dois a dois distintos em um corpo  $\mathbb{K}$  (e.g.,  $\mathbb{R}$ ,  $\mathbb{C}$ ,  $\mathbb{F}_p$ ), e sejam  $y_0, \dots, y_{N-1} \in \mathbb{K}$ . Existe um **único** polinômio  $p(x) = a_0 + a_1x + \dots + a_{N-1}x^{N-1} \in \mathbb{K}[x]$  tal que  $p(x_i) = y_i$  para todo  $i = 0, \dots, N-1$ .*

*Proof.* Escreva  $p(x) = \sum_{k=0}^{N-1} a_k x^k$ . Impor as condições  $p(x_i) = y_i$  para  $i = 0, \dots, N-1$  gera o sistema linear

$$\begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^{N-1} \\ 1 & x_1 & x_1^2 & \cdots & x_1^{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N-1} & x_{N-1}^2 & \cdots & x_{N-1}^{N-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-1} \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix}.$$

Denote essa matriz por  $V$  (matriz de Vandermonde, i.e.  $V_{i,j} = x_i^{j-1}$  para todo os índices  $i, j$ ), o vetor de coeficientes por  $a$  e o vetor de valores por  $y$ ; então o sistema é

$$Va = y.$$

O determinante de Vandermonde é dado por

$$\det(V) = \prod_{0 \leq i < j \leq N-1} (x_j - x_i).$$

Como os  $x_i$  são dois a dois distintos, temos  $x_j - x_i \neq 0$  para  $i \neq j$ , logo  $\det(V) \neq 0$ . Portanto,  $V$  é invertível e o sistema  $Va = y$  tem **solução única**, dada por

$$a = V^{-1}y.$$

Concluimos que existe um único vetor de coeficientes  $(a_0, \dots, a_{N-1})$ , isto é, um **único** polinômio  $p(x)$  de grau  $\leq N - 1$  que interpola os  $N$  pontos.  $\square$