

Escola Politécnica-USP
LARC



Seminário NTT

Resumo

O objetivo do semanário é apresentar a NTT e suas vantagens, para isso, parte-se da transformada de Fourier, em seguida, para a sua forma discreta (DFT) e chega-se na NTT propriamente dita. O que almejo com esse trabalho é passar a intuição por trás das ferramentas matemática utilizadas e beleza inerentes a elas.

Contents

1	A Dualidade entre Tempo e Frequênci a	2
2	Transformada de Fourier Contínua (CTFT)	2
3	Transformada Discreta de Fourier (DFT)	2
3.1	Definição da IDFT	3
4	A Multiplicação de Polinômios e a Complexidade Computacional	4
5	A Fast Fourier Transform	5
6	Problemas da FFT	8
7	Number Theoretic Transform (NTT)	8
7.1	Fundamentos	8
8	O Isomorfismo via Teorema Chinês dos Restos (CRT)[1]	11
9	Transformada Numérica Inversa (INTT)	12
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
9.3	Extra: NTT incompleta e a restrição de parâmetros no caso negacíclico	18
10	Agradecimentos	19
11	Apêndice	19
11.1	DFT como transformação linear	19
11.2	Exemplo de convolução negacíclica	20
11.3	Determinição unívoca do polinomio	22

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:

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

Elá pode ser entendida como um produto interno (uma projeção) do sinal com todos as frequencias da reta real $\langle g, e^{-2\pi ift} \rangle$, que, devido a ortogonalidade das frequências diferentes e que funcoes bem comportadas podem ser decompostas em series de autofuncoes $e^{-2\pi ift}$, consegue extrair extamente 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** (*Discrete Fourier Transform*). Ela opera sobre uma sequência finita de N amostras, mapeando dados discretos no tempo para o domínio da frequência. Segundo o Teorema de Amostragem de Nyquist-Shannon [2], para que um sinal contínuo possa ser unicamente determinado a partir de suas amostras, a frequência de amostragem f_s deve ser superior ao dobro da maior

frequência f_{max} contida no sinal. $f_s > 2f_{max}$. Com essa restrição para a quantidade de amostras N em mente, 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 11.1, 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:

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

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:

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

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

Além disso, a DFT pode 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. Define

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

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 $A_n \equiv X[n]$, oportamente, 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 determinação unívoca do polinômio de grau $N - 1$ por N pontos veja a seção 11.3 para demonstração desse fato)

4 A Multiplicação de Polinômios e a Complexidade Computacional

Um problema simplificado pela mudança de domínio é a multiplicação de polinomios. Tome os polinomios $f(x)$ 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 negacíclica** (de comprimento N), definida por:

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

Diferente da circular, aqui os termos que “ultrapassam” o comprimento N retornam com sinal invertido.

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:

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

Assim, o cálculo custoso da convolução é convertido em um produto ponto a ponto, pois a base de Fourier apropriada (utilizando raízes de $x^N + 1$) diagonaliza o operador de convolução negacíclica (matriz negacirculante). Com a transformada direta ingênuo o custo ainda seria $O(N^2)$, portanto, não há ganho computacional em realizar essa transformação.

Um exemplo para o convencimento do leitor foi disposto no apêndice 11.2.

5 A Fast Fourier Transform

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

O algoritmo da FFT foi redescoberto 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

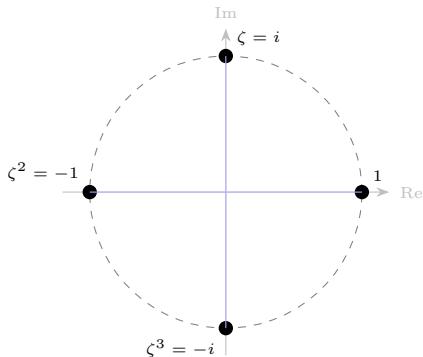
As raízes unitárias possuem propriedades cíclicas e certas simetrias que permitem a economia nos cálculos, vejamos um exemplo.

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

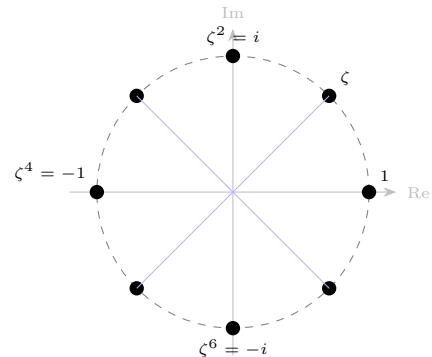
$$\zeta^1 = -i \quad \zeta^2 = -1$$

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

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 $\zeta_N^{a+N} = \zeta_N^a$ e $\zeta_N^2 = \zeta_{N/2}$, esta última pode ser vista na figura 1

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:

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

Usando $\zeta_N^2 = \zeta_{N/2}$:

$$\zeta_N^{2mk} = (\zeta_N^2)^{mk} = \zeta_{N/2}^{mk}$$

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

$$= E[k] + \zeta_N^k O[k], \quad k = 0, \dots, \frac{N}{2} - 1.$$

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

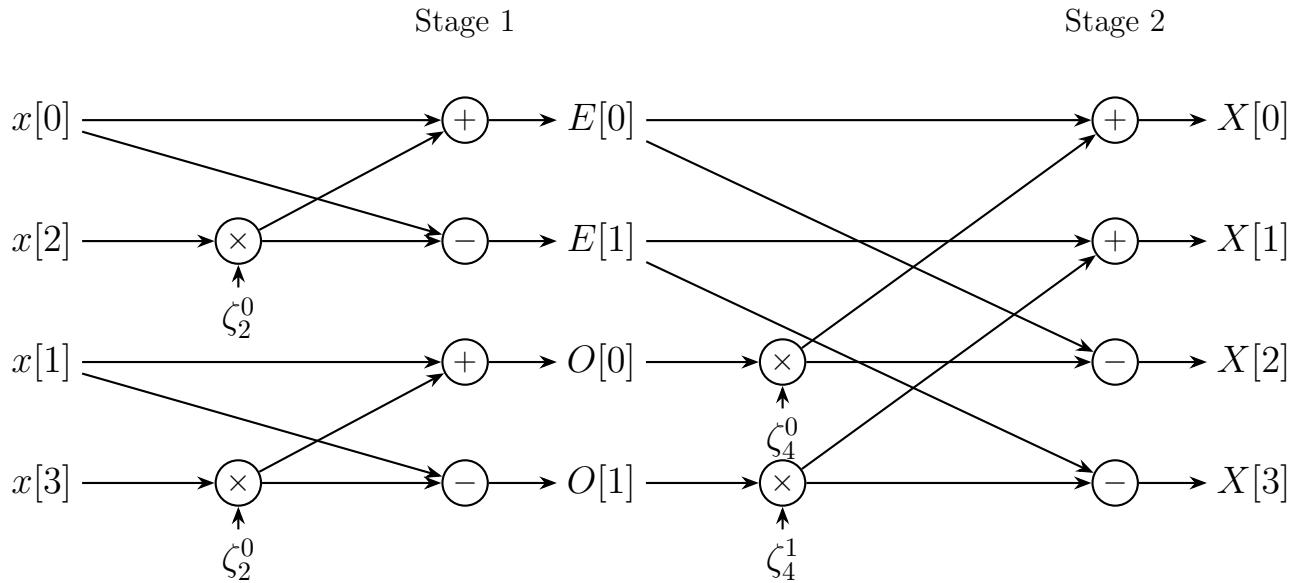


Figure 2: butterfly radix-2

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
19    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 convolução**, para multiplicar polinomios em $O(n \cdot \log n)$

6 Problemas da FFT

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

Solução: utilizar uma transformada que utiliza apenas números 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 ζ_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 DFT 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 ζ_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 ψ . 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 operações visualmente.)

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

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

Na forma matricial:

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

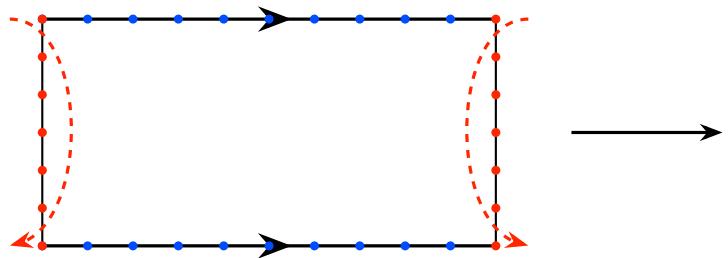
Neste caso, a matriz de transformação \mathbf{W}_N difere da versão cíclica padrão, pois seus coeficientes seguem a estrutura das raízes negacíclicas:

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

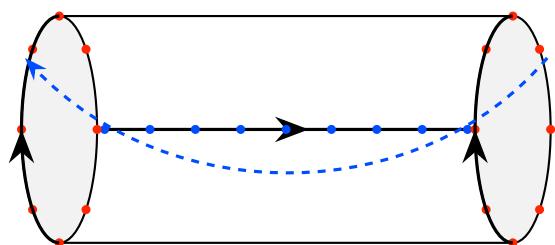
De forma geral, o termo na linha k e coluna n da matriz é dado por:

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

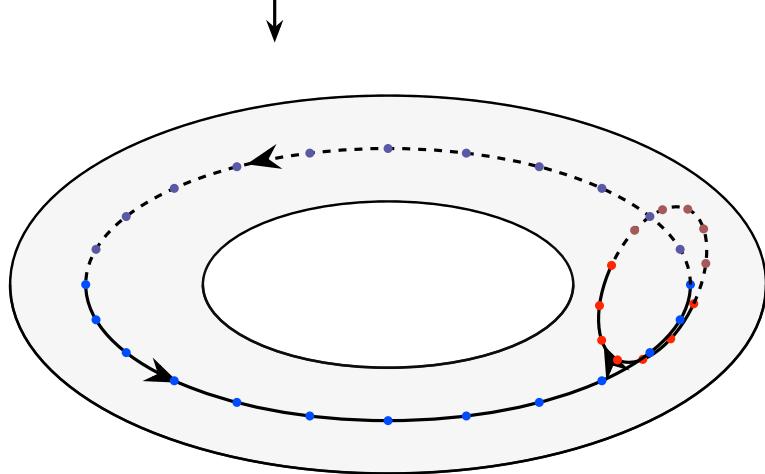
1. Colar lados horizontais (Passo \mathbb{Z}_p)



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 transformações

8 O Isomorfismo via Teorema Chinês dos Restos (CRT)[1]

A fundamentação algébrica da NTT reside na estrutura do anel quociente $\mathbb{Z}_p[x]/(x^N + 1)$. Já que $x^N + 1$ fatora-se em N fatores lineares distintos e coprimos,

$$x^N + 1 = \prod_{k=0}^{N-1} (x - \psi^{2k+1}),$$

Homomorfismo de avaliação e o isomorfismo local. Para cada $\alpha \in \mathbb{Z}_p$, considere o mapa de avaliação

$$\text{ev}_\alpha : \mathbb{Z}_p[x] \rightarrow \mathbb{Z}_p, \quad \text{ev}_\alpha(a(x)) = a(\alpha).$$

Esse mapa é um *homomorfismo de anéis* (preserva soma e produto):

$$\text{ev}_\alpha(a + b) = \text{ev}_\alpha(a) + \text{ev}_\alpha(b), \quad \text{ev}_\alpha(ab) = \text{ev}_\alpha(a) \text{ev}_\alpha(b).$$

Além disso, seu núcleo é exatamente o ideal gerado por $(x - \alpha)$:

$$\ker(\text{ev}_\alpha) = (x - \alpha).$$

De fato, pela Divisão Euclidiana, qualquer $a(x)$ pode ser escrito como

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

com $r \in \mathbb{Z}_p$, e avaliando em $x = \alpha$ obtemos $a(\alpha) = r$. Logo $a(\alpha) = 0$ se e somente se $(x - \alpha) \mid a(x)$. Pelo **Primeiro Teorema do Isomorfismo** para anéis,

$$\frac{\mathbb{Z}_p[x]}{(x - \alpha)} \cong \text{Im}(\text{ev}_\alpha) = \mathbb{Z}_p, \quad [a(x)] \longmapsto a(\alpha).$$

O isomorfismo global do CRT como produto de avaliações. Como os fatores $(x - \psi^{2k+1})$ são coprimos dois a dois, o CRT fornece um isomorfismo de anéis

$$\Phi : \frac{\mathbb{Z}_p[x]}{(x^N + 1)} \xrightarrow{\cong} \prod_{k=0}^{N-1} \frac{\mathbb{Z}_p[x]}{(x - \psi^{2k+1})} \cong \underbrace{\mathbb{Z}_p \times \cdots \times \mathbb{Z}_p}_{N \text{ vezes}}.$$

E, crucialmente, Φ é construído a partir dos homomorfismos de avaliação: a k -ésima componente é exatamente o homomorfismo $\text{ev}_{\psi^{2k+1}}$ (passando ao quociente). Assim, para $[a(x)] \in \mathbb{Z}_p[x]/(x^N + 1)$,

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

Multiplicação vira produto ponto a ponto. Como cada $\text{ev}_{\psi^{2k+1}}$ é homomorfismo e Φ é isomorfismo de anéis,

$$\Phi([a(x)][b(x)]) = \Phi([a(x)]) \cdot \Phi([b(x)])$$

onde o produto do lado direito é componente a componente. Em particular, se $C = \Phi([a(x)b(x)])$, então

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

Isso explica formalmente por que a multiplicação no quociente $\mathbb{Z}_p[x]/(x^N + 1)$ (convolução negacíclica dos coeficientes) se traduz em um produto de Hadamard no domínio transformado.

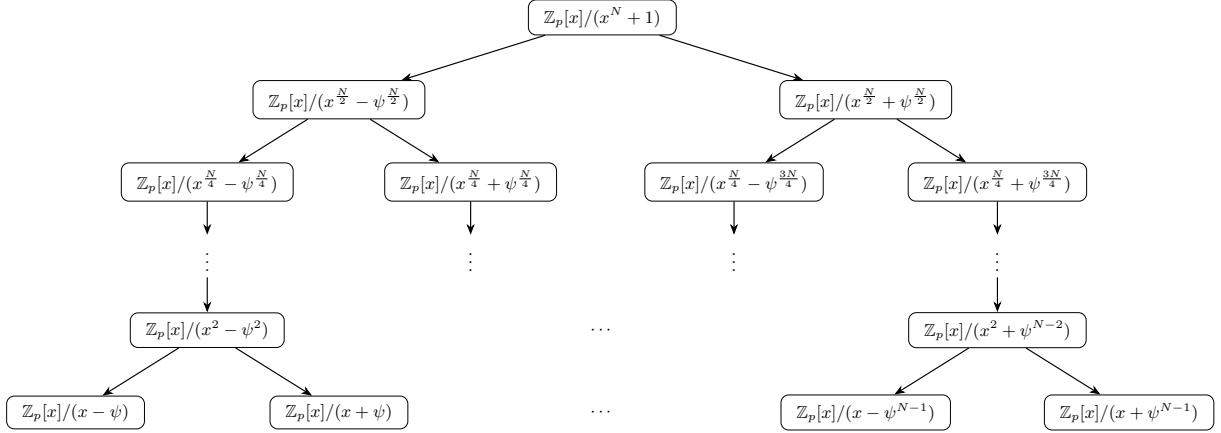


Figure 4: CRT da NTT Negacíclica

9 Transformada Numérica Inversa (INTT)

Como o Teorema Chinês dos Restos garante que a aplicação da NTT é um isomorfismo bijutor 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 $A(x)$ 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 ψ é uma raiz primitiva $2N$ -ésima da unidade, temos que $\psi^N \equiv -1 \pmod{p}$, o que implica $(\psi^{N/2})^2 \equiv -1$.

Pelo Teorema Chinês dos Restos (CRT), podemos fatorar o módulo original em dois sub-anéis:

$$x^N + 1 = x^N - (\psi^{N/2})^2 = \underbrace{(x^{N/2} - \psi^{N/2})}_{\text{Módulo Esquerdo}} \cdot \underbrace{(x^{N/2} + \psi^{N/2})}_{\text{Módulo Direito}}$$

Nota: Essa fatoração particiona o conjunto das raízes de $x^N + 1$ (que são as potências ímpares ψ^{2k+1}). As raízes que satisfazem o lado esquerdo continuam na recursão L, e as que satisfazem o lado direito seguem na recursão R.

2. O Colapso da Variável

Considere o polinômio de entrada $A(x)$ com coeficientes $a[i]$. Podemos dividir seu somatório em duas partes: os primeiros $N/2$ termos e os últimos $N/2$ termos.

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

Colocando $x^{N/2}$ em evidência na segunda metade, obtemos uma estrutura que revela a simetria do problema:

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

A intuição chave da NTT é projetar essa equação nos sub-anéis definidos no passo 1. Ao fazer isso, a potência $x^{N/2}$ deixa de ser uma variável e torna-se uma constante escalar:

- **Ramo Esquerdo** ($x^{N/2} \equiv \psi^{N/2}$): Substituímos $x^{N/2}$ por $+\psi^{N/2}$.
- **Ramo Direito** ($x^{N/2} \equiv -\psi^{N/2}$): Substituímos $x^{N/2}$ por $-\psi^{N/2}$.

3. A Operação Borboleta (Butterfly)

Ao realizar as substituições acima diretamente no somatório, os coeficientes dos novos polinômios reduzidos (a_L e a_R) surgem imediatamente. Para cada posição $0 \leq i < N/2$:

$$\begin{aligned} a_L[i] &= a[i] + \psi^{N/2} \cdot a[i+N/2] \pmod{p} \\ a_R[i] &= a[i] - \psi^{N/2} \cdot a[i+N/2] \pmod{p} \end{aligned}$$

Note que não precisamos calcular polinômios inteiros; a operação acontece pontualmente nos coeficientes. Os valores a_L tornam-se a entrada para a recursão da esquerda, e a_R para a recursão da direita.

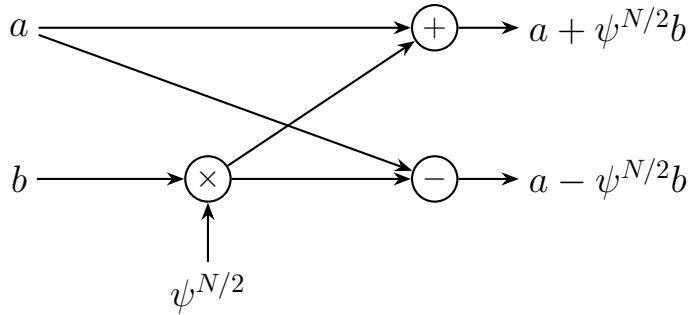


Figure 5: Butterfly NTT

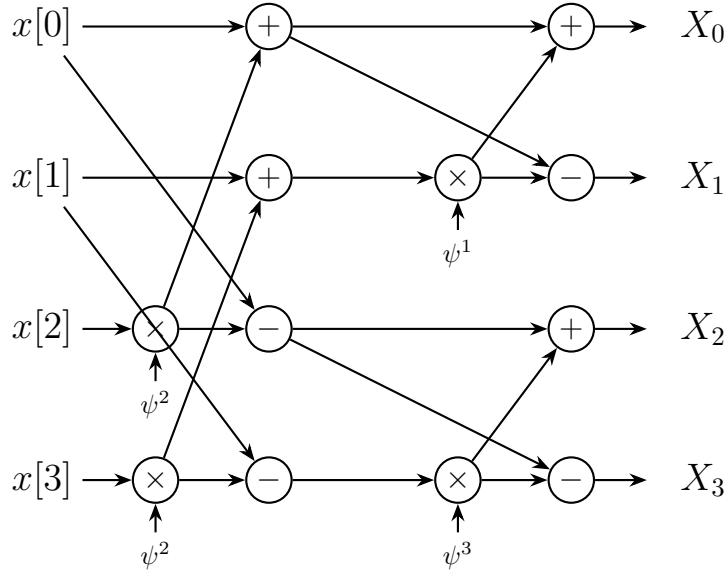


Figure 6: Butterfly NTT

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

Enquanto a transformação direta (NTT) utiliza a decomposição para reduzir o problema, a transformação inversa (INTT) realiza o caminho oposto: ela combina os resultados dos sub-problemas menores para reconstruir o vetor original. O algoritmo de Gentleman-Sande implementa essa reconstrução de forma eficiente.

1. A Inversão do Sistema Linear

Na etapa da transformação direta, definimos os valores dos sub-ramos (a_L e a_R) através das equações:

$$\begin{cases} a_L[i] = a[i] + \psi^{N/2} \cdot a[i + N/2] \\ a_R[i] = a[i] - \psi^{N/2} \cdot a[i + N/2] \end{cases}$$

Na etapa inversa, conhecemos a_L e a_R (retornados pela recursão) e nosso objetivo é recuperar os coeficientes originais $a[i]$ e $a[i + N/2]$. Trata-se apenas de resolver este sistema linear de duas variáveis.

2. A Lógica da Recuperação

Somando e subtraindo as duas equações do sistema acima, podemos isolar os termos desejados:

- **Recuperando a metade inferior ($a[i]$):** Somamos as equações:

$$a_L[i] + a_R[i] = 2a[i] \implies a[i] = \frac{a_L[i] + a_R[i]}{2}$$

- **Recuperando a metade superior ($a[i + N/2]$):** Subtraímos a segunda da primeira:

$$a_L[i] - a_R[i] = 2\psi^{N/2} \cdot a[i + N/2]$$

Para isolar $a[i + N/2]$, multiplicamos pelo inverso de $2\psi^{N/2}$:

$$a[i + N/2] = \frac{a_L[i] - a_R[i]}{2} \cdot \psi^{-N/2}$$

Esta manipulação algébrica revela a diferença fundamental na ordem das operações entre a ida (Cooley-Tukey) e a volta (Gentleman-Sande).

3. A Borboleta GS (Inverse Butterfly)

Baseado nas isolações acima, definimos a operação borboleta inversa. Para cada par de entradas $a_L[i]$ e $a_R[i]$ provenientes dos sub-blocos, calculamos:

$$a[i] = \frac{1}{2}(a_L[i] + a_R[i]) \pmod{p}$$

$$a[i + N/2] = \frac{1}{2}(a_L[i] - a_R[i]) \cdot \psi^{-N/2} \pmod{p}$$

Observe a mudança estrutural:

1. Na **Ida (CT)**, multiplicamos pelo fator de rotação *antes* de somar/subtrair.
2. Na **Volta (GS)**, subtraímos primeiro e multiplicamos pelo fator de rotação inverso *depois*.

Nota de Implementação: O fator escalar $1/2$ (inverso modular de 2) geralmente não é aplicado a cada camada. Para eficiência, acumula-se o fator total $1/N$ e aplica-se uma única multiplicação escalar ao final de toda a transformação.

4. Fluxo de Dados e Bit-Reversal

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* (permutada), a INTT via GS aceita os dados em ordem *bit-reversed* e os reconstrói naturalmente para a ordem linear.

Isso cria um par perfeito:

$$\text{Input Natural} \xrightarrow{\text{CT}} \text{Bit-Reversed} \xrightarrow{\text{Operação Pontual}} \text{Bit-Reversed} \xrightarrow{\text{GS}} \text{Output Natural}$$

Essa simetria elimina a necessidade de reordenamentos de memória (bit-reversal) custosos entre as transformações.

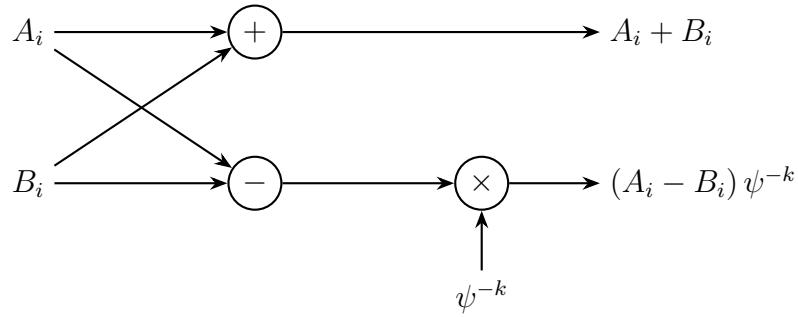


Figure 7: Butterfly INTT

A implementa da NTT negacíclica ésta 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 comparação 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

fonte: <https://github.com/SheafificationOfG/Fibsonisheaf>

9.3 Extra: NTT incompleta e a restrição de parâmetros no caso negacíclico

Na multiplicação negacíclica de polinômios no anel

$$\mathbb{Z}_p[x]/(x^N + 1),$$

com N potência de dois, a utilização de uma **NTT completa** exige a existência de uma raiz primitiva de ordem $2N$ módulo p . Quando p é primo, essa condição é equivalente a

$$p \equiv 1 \pmod{2N}.$$

Em aplicações criptográficas baseadas em reticulados, em especial em esquemas de Fully Homomorphic Encryption o valor de N é tipicamente muito grande. Como consequência, a condição acima impõe fortes restrições sobre a escolha do primo p , frequentemente forçando o uso de módulos grandes ou pouco flexíveis, o que impacta tanto a eficiência quanto o ajuste fino de parâmetros de segurança.

Uma forma natural de relaxar essa restrição é empregar a chamada **NTT ℓ -incompleta**. A ideia consiste em interromper o algoritmo da NTT antes de executar todos os $\log_2 N$ estágios do esquema radix-2. Mais precisamente, ao parar após $\log_2 N - \ell$ estágios, a existência da transformada passa a requerer apenas uma raiz da unidade de ordem $2N/2^\ell$ em \mathbb{Z}_p , o que resulta na condição mais fraca

$$p \equiv 1 \pmod{2N/2^\ell}.$$

Dessa forma, o conjunto de primos admissíveis torna-se significativamente maior, permitindo escolhas de parâmetros mais flexíveis.

Do ponto de vista algébrico, a NTT ℓ -incompleta não avalia o polinômio em todas as N raízes da unidade, mas o mapeia para um vetor de $N/2^\ell$ polinômios de menor grau, cada um pertencente a um quociente do tipo

$$\mathbb{Z}_p[x]/(x^{2^\ell} - \psi_i),$$

onde ψ_i são potências apropriadas da raiz disponível. A multiplicação passa então a ser realizada componente a componente nesses anéis menores, seguida de uma transformada inversa incompleta.

O principal custo adicional desse método está na *multiplicação de base* dentro dos anéis $\mathbb{Z}_p[x]/(x^{2^\ell} - \psi_i)$, que, para valores maiores de ℓ , tende a ser implementada por algoritmos quadráticos. No entanto, o trabalho de Paiva et al. [3] mostra que, no contexto de bootstrapping amortizado para esquemas do tipo FHEW/TFHE, esse custo pode ser eliminado. Os autores reformulam a NTT inversa como o produto de duas matrizes de borboleta (controladas por um parâmetro de divisão α) e demonstram que a multiplicação de base da NTT incompleta pode ser *incorporada* ao primeiro estágio dessa decomposição matricial, sem aumento no número assintótico de operações.

Como resultado, obtém-se o que os autores denominam “*NTT incompleta gratuita*”: para valores moderados de ℓ , é possível relaxar substancialmente a condição sobre o primo p sem penalidade computacional relevante, ao mesmo tempo em que se ampliam as opções de parâmetros e se melhora o compromisso entre desempenho e taxa de falha (decryption failure rate).

10 Agradecimentos

Gostaria de agradecer ao professor doutor *Thales Paiva* pela chance de apresentar o semanário e pelos ensinamentos, gostaria de agradecer também o professor doutor *João Fernando da Cunha Nariyoshi* por esclarecer algumas dúvidas sobre a parte da matemática pura. Foi realmente divertido pesquisar, escrever, desenhar e animar para esse semanário.

11 Apêndice

11.1 DFT como transformação linear

Seja ζ uma raiz primitiva N -ésima da unidade, por exemplo

$$\zeta = e^{-2\pi i/N}.$$

Dado um sinal discreto de comprimento N , com amostras $x[0], x[1], \dots, x[N-1]$ (no domínio do tempo), definimos a DFT como a sequência $X[0], X[1], \dots, X[N-1]$ (no domínio da frequência), onde

$$X[k] = \sum_{n=0}^{N-1} x[n] \zeta^{kn}, \quad k = 0, 1, \dots, N-1.$$

Prova de linearidade

Sejam dois sinais $x[n]$ e $y[n]$ e escalares $a, b \in \mathbb{C}$. Considere a combinação linear

$$z[n] = a x[n] + b y[n].$$

Seja $Z[k]$ a DFT de $z[n]$. Então

$$\begin{aligned} Z[k] &= \sum_{n=0}^{N-1} z[n] \zeta^{kn} = \sum_{n=0}^{N-1} (a x[n] + b y[n]) \zeta^{kn} \\ &= a \sum_{n=0}^{N-1} x[n] \zeta^{kn} + b \sum_{n=0}^{N-1} y[n] \zeta^{kn} = a X[k] + b Y[k]. \end{aligned}$$

Logo,

$$\text{DFT}(ax + by) = a \text{DFT}(x) + b \text{DFT}(y),$$

o que mostra que a DFT é uma **transformação linear**.

11.2 Exemplo de convolução negacíclica

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

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

Neste contexto, a frequência fundamental não é a raiz da unidade, mas sim a raiz de $x^N + 1 = 0$. Seja ψ tal que $\psi^N = -1$. Para $N = 3$, temos $\psi^3 = -1$.

Método 1: Convolução no Tempo (Matriz Negacirculante)

A convolução negacíclica $z = x *_{\text{neg}} y$ equivale à multiplicação de uma matriz negacirculante H_x pelo vetor y . Note que os elementos que “dão a volta” (acima da diagonal principal) têm sinal invertido:

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

Calculando $z = H_x y$:

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

Método 2: Diagonalização (Domínio da Frequência)

A matriz negacirculante H_x é diagonalizada por uma variante da DFT que utiliza potências ímpares de ψ (raízes de $x^N + 1$). Matematicamente:

$$H_x = \mathcal{F}^{-1} \Lambda_x \mathcal{F}$$

onde Λ_x contém os autovalores de H_x , que correspondem à avaliação do polinômio $x(z)$ nas raízes de $z^N = -1$.

Para verificar os autovalores (o espectro de x), calculamos o polinômio característico de H_x :

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

Expandindo o determinante (regra de Sarrus):

$$(1-\lambda)^3 - (2)(2)(2) = 0 \implies (1-\lambda)^3 = 8$$

As raízes para $(1-\lambda)$ são as raízes cúbicas de 8. Sabemos que $8 = 8 \cdot 1$, mas no contexto complexo as raízes são $2, 2\zeta_3$ e $2\zeta_3^2$. Logo:

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

Estes valores λ_k correspondem à Transformada Negacíclica (NTT sobre $x^N + 1$) do vetor x , permitindo calcular a convolução via produto ponto a ponto.

Verificação do Autovetor

Verificamos agora se o autovetor v_1 da base Negacíclica (coluna de \mathcal{F}^{-1} associada à raiz ψ), dado por $v_1 = [1, \psi^{-1}, \psi^{-2}]^T = [1, -\psi^2, -\psi]^T$, satisfaz $H_x v_1 = \lambda_1 v_1$.

Lado esquerdo ($H_x v_1$):

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

Lado direito ($\lambda_1 v_1$), onde o autovalor é $\lambda_1 = x(\psi) = 1 + 2\psi$:

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

Utilizando a propriedade $\psi^3 = -1$, simplificamos o termo do meio: $-\psi^2 - 2(-1) = 2 - \psi^2$. A igualdade é satisfeita, confirmando que as raízes de $x^N + 1$ geram a base natural de H_x .

Representação Matricial da Diagonalização

Primeiro, calculamos os vetores transformados $X = \text{NTT}(x)$ e $Y = \text{NTT}(y)$ avaliando os polinômios nas raízes de $z^3 = -1$:

$$X = \begin{bmatrix} x(-1) \\ x(\psi) \\ x(\psi^2) \end{bmatrix} = \begin{bmatrix} -1 \\ 1+2\psi \\ 1+2\psi^2 \end{bmatrix}, \quad Y = \begin{bmatrix} y(-1) \\ y(\psi) \\ y(\psi^2) \end{bmatrix} = \begin{bmatrix} 2 \\ 1+\psi^2 \\ 1+\psi^4 \end{bmatrix}$$

*Nota: $y(\psi) = 1 + 0\psi + 1\psi^2$. Para o terceiro termo, usamos $\psi^4 = -\psi$, logo $1 - \psi$.

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

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

Executando o produto ponto a ponto:

$$\mathbf{Z} = \begin{bmatrix} -2 \\ (1+2\psi)(1+\psi^2) \\ (1+2\psi^2)(1-\psi) \end{bmatrix} = \begin{bmatrix} -2 \\ 1+\psi^2+2\psi+2\psi^3 \\ 1-\psi+2\psi^2-2\psi^3 \end{bmatrix} = \begin{bmatrix} -2 \\ -1+2\psi+\psi^2 \\ 3-\psi+2\psi^2 \end{bmatrix}$$

*Simplificações usando $\psi^3 = -1$.

Retorno ao Tempo (INTT)

Finalmente, aplicamos a transformada inversa. O vetor resultante z deve coincidir com o cálculo temporal $[-1, 2, 1]^T$. Ao reconstruir o polinômio $z(\omega)$ a partir dos valores em \mathbf{Z} , obtemos os coeficientes:

$$z = \text{INTT}(\mathbf{Z}) = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$$

Isso verifica que:

$$z(\psi) = -1 + 2\psi + \psi^2$$

O que coincide exatamente com o segundo elemento do vetor \mathbf{Z} calculado acima, fechando o ciclo da prova numérica.

11.3 Determinação unívoca do polinomio

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

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

Concluímos 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