NTT for Negacyclic Polynomial Multiplication

arnaucube

January 2025

Abstract

Notes taken while studying the NTT, mostly from [1].

Usually while reading books and papers I take handwritten notes in a notebook, this document contains some of them re-written to $LaTeX$.

The notes are not complete, don't include all the steps neither all the proofs.

An implementation of the NTT can be found at https://github.com/arnaucube/fhe-study/blob/main/arithmetic/src/ntt.rs.

Contents

1 Main idea

For doing multiplications in the *negacyclic polynomial ring* ($\mathbb{Z}_q[X]/(X^n + 1)$), rather than doing it in a naive way, it is more efficient to do it through the NTT.

This is, let $a(X), b(X) \in \mathbb{Z}_q[X]/(X^n + 1)$, and suppose we want to obtain $a(X) \cdot b(X)$. First apply the NTT to the two ring elements that we want to multiply,

$$\hat{a}(X) = NTT(a(X)), \quad \hat{b}(X) = NTT(b(X))$$

then multiply the result element-wise,

$$c = \hat{a} \circ \hat{b}$$

where $\circ$ means the element-wise vector multiplication in $\mathbb{Z}_q$.

Then apply the NTT^{-1} to the result, obtaining the actual value of multiplying $a(X) \cdot b(X)$.

2 Cyclotomic vs Negacyclic

2.1 Cyclotomic: $\mathbb{Z}_q[X]/(X^n - 1)$

In the cyclotomic case, the primitive n-th root of unity in Z_q is $w^n \equiv 1 \pmod q$ (and $w^k \not\equiv 1 \pmod q \quad fork < n$)

2.1.1 NTT based on w

NTT of a polynomial $a(X) = \sum a_i X^i$ is defined as $\hat{a} = NTT(a)$, where

$$\hat{a}_j = \sum_{i=0}^{n-1} a_i w^{ij} \pmod q$$

for each of the $j = 0, 1, \ldots, n-1$.

We can visualize the NTT operation as

$$NTT(a) = \begin{bmatrix} w^{0\cdot 0} & w^{0\cdot 1} & w^{0\cdot 2} & \ldots & w^{0\cdot(n-1)} \\ w^{1\cdot 0} & w^{1\cdot 1} & w^{1\cdot 2} & \ldots & w^{1\cdot(n-1)} \\ w^{2\cdot 0} & w^{2\cdot 1} & w^{2\cdot 2} & \ldots & w^{2\cdot(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ w^{(n-1)\cdot 0} & w^{(n-1)\cdot 1} & w^{(n-1)\cdot 2} & \ldots & w^{(n-1)\cdot(n-1)} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_{n-1} \end{bmatrix}$$

2.1.2 Inverse NTT based on w

Inverse-NTT of a vector $\hat{a}$ is defined as $a = iNTT(\hat{a})$, where

$$a_i = n^{-1} \sum_{j=0}^{n-1} \hat{a}_j w^{-ij} k \pmod q$$

with $j = 0, 1, \ldots, n-1$.

Similar to the NTT formula, only diffs:

- w is replaced by its inverse in $\mathbb{Z}_q$

- n^{-1} scaling factor

We can visualize the NTT^{-1} operation as

$$iNTT(\hat{a}) = n^{-1} \cdot \begin{bmatrix} w^{-0\cdot 0} & w^{-0\cdot 1} & w^{-0\cdot 2} & \dots & w^{-0\cdot(n-1)} \\ w^{-1\cdot 0} & w^{-1\cdot 1} & w^{-1\cdot 2} & \dots & w^{-1\cdot(n-1)} \\ w^{-2\cdot 0} & w^{-2\cdot 1} & w^{-2\cdot 2} & \dots & w^{-2\cdot(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ w^{-(n-1)\cdot 0} & w^{-(n-1)\cdot 1} & w^{-(n-1)\cdot 2} & \dots & w^{-(n-1)\cdot(n-1)} \end{bmatrix} \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_{n-1} \end{bmatrix} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{n-1} \end{bmatrix}$$

2.2 Apply it to polynomial multiplication

Want to compute $c(X) = a(X) \cdot b(X) \in \mathbb{Z}_q[X]/(X^n - 1)$, which we can do as

$$c = iNTT(NTT(a) \circ NTT(b))$$

where $\circ$ means the element-wise vector multiplication in $\mathbb{Z}_q$.

2.3 Negacyclic: $\mathbb{Z}_q[X]/(X^n + 1)$

Instead of working in $\mathbb{Z}_q[X]/(X^n - 1)$, we work in $\mathbb{Z}_q[X]/(X^n + 1)$.

Instead of using the primitive n-th root of unity (w), we use the *primitive 2n-th root of unity* ψ.

Where $\psi^2 \equiv w \pmod{q}$, and $\psi^2 \equiv -1 \pmod{q}$.

2.3.1 NTT based on ψ, NTT$^\psi$

$\hat{a} = NTT^\psi(a)$, where

$$\hat{a}_j = \sum_{i=0}^{n-1} \psi^i w^{ij} a_i \pmod{q}$$

with $j = 0, 1, \dots, n-1$.

Since $\psi^2 \equiv w \pmod{q}$, we can substitute $w = \psi^2$:

$$\hat{a}_j = \sum_{i=0}^{n-1} \psi^{2ij+i} a_i \pmod{q}$$

getting rid of w.

We can visualize the NTT$^\psi$ operation as

$$NTT^\psi(a) = \begin{bmatrix} \psi^{2(0\cdot 0)+0} & \psi^{2(0\cdot 1)+1} & \psi^{2(0\cdot 2)+2} & \dots & \psi^{2(0\cdot(n-1))+(n-1)} \\ \psi^{2(1\cdot 0)+0} & \psi^{2(1\cdot 1)+1} & \psi^{2(1\cdot 2)+2} & \dots & \psi^{2(1\cdot(n-1))+(n-1)} \\ \psi^{2(2\cdot 0)+0} & \psi^{2(2\cdot 1)+1} & \psi^{2(2\cdot 2)+2} & \dots & \psi^{2(2\cdot(n-1))+(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ \psi^{2((n-1)\cdot 0)+0} & \psi^{2((n-1)\cdot 1)+1} & \psi^{2((n-1)\cdot 2)+2} & \dots & \psi^{2((n-1)\cdot(n-1))+(n-1)} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_{n-1} \end{bmatrix}$$

2.3.2 Inverse NTT based on ψ, iNTT$^{\psi}$

$a = iNTT\psi(\hat{a})$, where

$$a_i = n^{-1} \sum_{j=0}^{n-1} \psi^{-j} w^{-ij} \hat{a}_j \pmod{q}$$

with $i = 0, 1, \ldots, n-1$.

Which substituting $w = \psi^2$ we get

$$a_i = n^{-1} \sum_{j=0}^{n-1} \psi^{-(2ij+j)} \hat{a}_j \pmod{q}$$

So the differences with the NTT$^{\psi}$ are:

- ψ is replaced by its inverse ψ^{-1} in $\mathbb{Z}_q$
- n^{-1} scaling factor
- transpose of the exponents of ψ

We can visualize the NTT$^{-\psi}$ operation as

$iNTT^{\psi}(a) =$

$$\begin{bmatrix} \psi^{-(2(0\cdot0)+0)} & \psi^{-(2(0\cdot1)+1)} & \psi^{-(2(0\cdot2)+2)} & \ldots & \psi^{-(2(0\cdot(n-1))+(n-1))} \\ \psi^{-(2(1\cdot0)+0)} & \psi^{-(2(1\cdot1)+1)} & \psi^{-(2(1\cdot2)+2)} & \ldots & \psi^{-(2(1\cdot(n-1))+(n-1))} \\ \psi^{-(2(2\cdot0)+0)} & \psi^{-(2(2\cdot1)+1)} & \psi^{-(2(2\cdot2)+2)} & \ldots & \psi^{-(2(2\cdot(n-1))+(n-1))} \\ \vdots & \vdots & \vdots & & \vdots \\ \psi^{-(2((n-1)\cdot0)+0)} & \psi^{-(2((n-1)\cdot1)+1)} & \psi^{-(2((n-1)\cdot2)+2)} & \ldots & \psi^{-(2((n-1)\cdot(n-1))+(n-1))} \end{bmatrix} \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_{n-1} \end{bmatrix} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{n-1} \end{bmatrix}$$

2.4 Use it to polynomial multiplication

Want to compute $c(X) = a(X) \cdot b(X) \in \mathbb{Z}_q[X]/(X^n - 1)$, which we can do as

$$c = iNTT^{\psi 1}(NTT^{\psi}(a) \circ NTT^{\psi}(b))$$

where $\circ$ means the element-wise vector multiplication in $\mathbb{Z}_q$.

3 Fast NTT

NTT and INTT have $O(n^2)$ complexity, but since NTT is the DFT in a ring, we can apply the DFT optimization techniques (FFT), to reduce the complexity to $O(nlogn)$.

We use two properties of ψ:

- periodicity: $\psi^{k+2n} = \psi^k$
- symmetry: $\psi^{k+n} = -\psi^k$

3.1 Cooley-Tukey algorithm (Fast NTT)

Recall,

$$\hat{a}_j = \sum_{i=0}^{n-1} \psi^{2ij+i} a_i \pmod q$$

we can split it into two parts,

$$\begin{aligned}\hat{a}_j &= \sum_{i=0}^{n/2-1} \psi^{4ij+2i} a_{2i} + \sum_{i=0}^{n/2-1} \psi^{4ij+2j+2i+1} a_{2i+1} \pmod q \\ &= \sum_{i=0}^{n/2-1} \psi^{4ij+2i} a_{2i} + \psi^{2j+1} \cdot \sum_{i=0}^{n/2-1} \psi^{4ij+2i} a_{2i+1} \pmod q\end{aligned}$$

Let

$$\begin{aligned}A_j &= \sum_{i=0}^{n/2-1} \psi^{4ij+2i} a_{2i} \pmod q \\ B_j &= \sum_{i=0}^{n/2-1} \psi^{4ij+2i} a_{2i+1} \pmod q\end{aligned}$$

then,

$$\begin{aligned}\hat{a}_j &= A_j + \psi^{2j+1} \cdot B_j \pmod q \\ \hat{a}_{j+n/2} &= A_j - \psi^{2j+1} \cdot B_j \pmod q\end{aligned}$$

Notice that $A_j,\ B_j$ can be obtained as $n/2$ points. So if n is a power of two, we can repeat the process for all the coefficients.

[todo: diagram and explain intuition]

3.2 Gentleman-Sande algorithm (Fast iNTT)

Instead of dividing the summation by its index parity, it is separated by the lower and upper half of the summation.

Similar to what we did in section 3.1, let's split the equation to compute a_i.

Recall that we had

$$a_i = n^{-1} \sum_{j=0}^{n-1} \psi^{-(2ij+j)} \hat{a}_j \pmod q$$

we can split it into two parts,

$$
\begin{aligned}
a_i &= n^{-1} \cdot \left[\sum_{j=0}^{n/2-1} \psi^{-(2i+1)j} \hat{a}_j + \sum_{j=0}^{n/2-1} \psi^{-(2i+1)(j+n/2)} \hat{a}_{j+n/2} \right] \pmod{q} \\
&= n^{-1} \cdot \psi^{-i} \cdot \left[\sum_{j=0}^{n/2-1} \psi^{-2ij} \hat{a}_j + \sum_{j=0}^{n/2-1} \psi^{-2i(j+n/2)} \hat{a}_{j+n/2} \right] \pmod{q}
\end{aligned}
$$

Based on the periodicity and symmetry of ψ^{-1}, leaving the n^{-1} factor out, for the even terms:

$$
\begin{aligned}
a_{2i} &= \psi^{-2i} \cdot \left[\sum_{j=0}^{n/2-1} \psi^{-4ij} \hat{a}_j + \sum_{j=0}^{n/2-1} \psi^{-4i(j+n/2)} \hat{a}_(j+n/2) \right] \pmod{q} \\
&= \psi^{-2i} \sum_{j=0}^{n/2-1} (\hat{a}_j + \hat{a}_{j+n/2}) \psi^{-4ij}) \pmod{q}
\end{aligned}
$$

Doing the same derivation for the odd terms:

$$
a_{2i+1} = \psi^{-2i} \sum_{j=0}^{n/2-1} (\hat{a}_j - \hat{a}_{j+n/2}) \psi^{-4ij} \pmod{q}
$$

Now, let

$$
A_j = \sum_{j=0}^{n/2-1} \hat{a}_j \psi^{-4ij}, \quad B_j = \sum_{j=0}^{n/2-1} \hat{a}_{j+n/2} \psi^{-4ij}
$$

then

$$
\begin{aligned}
a_{2i} &= (A_i + B_i) \psi^{-2i} \pmod{q} \\
a_{2i+1} &= (A_i - B_i) \psi^{-2i} \pmod{q}
\end{aligned}
$$

[todo: add diagram and explain intuition]

References

[1] Ardianto Satriawan, Infall Syafalni, Rella Mareta, Isa Anshori, Wervyan Shalannanda, and Aleams Barra. Conceptual review on number theoretic transform and comprehensive review on its implementations. *IEEE Access*, 11:70288–70316, 2023.