Universidad Autónoma Metropolitana

Unidad Iztapalapa

División de Ciencias Básicas e Ingeniería

ETRUSign: NTRUSign sobre los enteros de Eisenstein

Tesis que presenta:

Oscar Casimiro Muñoz

Para obtener el grado de:

Maestro en Ciencias
(Matemáticas Aplicadas e Industriales)

Asesor:

Dr. José Noé Gutíerrez Herrera

Ciudad de México, diciembre 2022

Agradecimientos

Agradezco hoy y siempre a mis padres, por brindarme su apoyo en todo momento, por ser la razón de todos mis logros, por no dejarme solo y por darme sus grandes consejos llenos de sabiduría.

Agradezco a mi asesor de tesis, el Dr. José Noé Gutiérrez Herrera, por guiarme en la elaboración de mi tesis, por todo su tiempo que me dedico para corregir mi tesis, por su paciencia que me tuvo para superar los temas que me eran difíciles.

Agradezco a mis sinodales, la Dra. Yuriko Pitones Amaro, el Dr. Juan Carlos Ku Cauich y el Dr. José Noé Gutiérrez Herrera, por sus sabias correcciones y consejos que me dieron para que esta tesis quedara de manera correcta.

Agradezco a CONACYT, por otorgarme una beca con número de CVU: 1082416, y así brindarme apoyo económico durante mis estudios de maestría.

Dedicatoria

A mis padres, lo más preciado que tengo en este mundo.

A la Universidad Autónoma Metropolitana, por ser una universidad muy generosa conmigo y con todos sus alumnos.

Índice general

Capítulo 1

Introducción

La criptografía estudia la manera de ocultar información bajo sistemas de cifrado o codificación en los cuales se altera la representación lingüística de ciertos mensajes con el fin de hacerlos ininteligibles a personas no autorizadas. Por lo tanto, el objetivo de la criptografía es conseguir la confidencialidad de la información. A medida que la economía mundial se vuelve más dependiente de la información en lugar de los bienes físicos, la criptografía se vuelve más esencial porque resguarda información valiosa, a la que solo se debe tener acceso controlado.

Los esquemas de firma digital son de gran utilidad para celebrar contratos de alguna tarjeta de crédito bancaria, certificar algún documento digital, realizar trámites de administración pública, etc., y se han vuelto indispensables para estas actividades cotidianas. Esta necesidad muy útil genera interés por estudiar diferentes tipos de esquemas de firma digital.

El objetivo principal de una firma digital es dar una versión análoga al propósito de una firma con pluma y tinta en un documento físico. De manera precisa si se supone que Aldo tiene un documento digital, M, y desea agregar información adicional que se pueda usar para probar posteriormente que Aldo aprobó o firmó el documento, entonces se puede ver la firma digital, $MSign$ de Aldo, como una analogía de su firma manuscrita con tinta sobre un papel físico.

Los reconocidos matemáticos Jeffrey Hoffstein, Jill Pipher y Joseph H. Silverman crearon un criptosistema llamado NTRU, el cual fue patentado en 2008 por sus grandes beneficios de velocidad para cifrar mensajes y por la seguridad que ofrece. Posteriormente estos mismos autores diseñaron un esquema de firma digital basado en otro esquema de firma digital llamado GGH. Este último, a su vez está basado en resolver el problema del vector más cercano (CVP) en un lattice, en donde las primeras versiones resultaron ser poco resistente a diversos ataques comunes contra los esquemas de firmas. En 2018 fue aprobada y estandarizada, por el Instituto Nacional de Estándares y Tecnología (NIST), una versión de esquema de firma llamada $pqNTRUSign$ que satisface ser resistente a al menos dos ataques comunes contra este esquema de firma: los ataques mediante lattices y los ataques de transcripción.

El objetivo general de esta tesis es brindar un esquema de firma digital análogo al esquema de firma estándar $pqNTRUSign$, pero que este implementado sobre el anillo de los enteros de Eisenstein y que conserve propiedades de seguridad que sean resistentes a los ataques comunes. Se proporciona un ejemplo del funcionamiento de este esquema de firma programado en SageMath 9.0.

Esta tesis consiste de 8 capítulos. En el capítulo 1 se redacta la introducción, la cual presenta el por qué se ha vuelto importante el estudio de esquemas de firmas digitales, algunos antecedentes y el objetivo general de la presente tesis. Después introducimos, en el capítulo 2, el concepto de lattice y algunas de sus propiedades, así como el funcionamiento del algoritmo LLL. Estos conceptos son de vital importancia para el estudio de esquemas de firma digital que se basan en lattices, y tales conceptos son estudiados en [20] y [24].

En el capítulo 3 se desarrolla el funcionamiento del criptosistema NTRU con una notación moderna introducida por J. Hoffstein, J. Pipher y J. H. Silverman en [10]. Se estudia la probabilidad de fallo del descifrado NTRU, así como un análisis de seguridad que consiste de un ataque mediante lattices presentado en [20].

El capítulo 4 describe las principales propiedades del anillo de los enteros de Eisenstein $\mathbb{Z}[\omega]$, por ejemplo, se prueba que $\mathbb{Z}[\omega]$ es un anillo euclidiano, se introduce un algoritmo de la división en este anillo, propuesto por Katherine Jarvis en [20], así como algunas relaciones entre números primos en este anillo y el anillo de los enteros $\mathbb{Z}$. Estos conceptos serán utilizados para formar un anillo base en la implementación el esquema de firma $ETRUSign$.

Se estudia, en el capítulo 5, el funcionamiento del criptosistema ETRU, diseñado por Katherine Jarvis en [20], el cálculo de la probabilidad de fallo de descifrado y un ataque de lattices a este criptosistema.

En el capítulo 6 se proporciona el concepto de esquema de firma digital y su funcionamiento con funciones hash,[13]. Se dan ejemplos del esquema de firma digital con el criptosistema RSA para motivar la implementación de un esquema de firma digital. También, en este mismo capítulo presentamos el esquema de firma digital $NTRUSign$ que está implementado con el criptosistema NTRU propuesto en [11], el cual es una versión similar al esquema de firma digital estándar $pqNTRUSign$, con propiedades de seguridad similares.

Se discute, en el capítulo 7, un nuevo esquema de firma digital llamado $ETRUSign$, que tiene un funcionamiento similar al esquema $NTRUSign$; se realiza un análisis de seguridad, y se compara la seguridad de este esquema con la del esquema $NTRUSign$, presentado en [15], el cual resulta ser de mayor seguridad ya que amplia considerablemente el espacio de firmas sin perder el mecanismo del esquema estándar $pqNTRUSign$ y el esquema $NTRUSign$; se proporciona un ejemplo del funcionamiento del esquema $ETRUSign$ programado en SageMath versión 9.0.

Finalmente, en el capítulo 8, se presentan las conclusiones y perspectivas de esta tesis.

Capítulo 2

Lattices y el algoritmo LLL

En este capítulo se presenta el concepto de lattice (también conocido como retícula, por su traducción al español) y algunas de sus propiedades. Un problema importante y difícil de resolver es encontrar el vector de menor longitud en un lattice, estudiado en [23]. Se estudiará el algoritmo LLL, que se utiliza para dar una solución eficiente al problema del vector más corto.

2.1. Definiciones y propiedades de lattices

El material de este capítulo aparece como conceptos estándar y se puede encontrar, por ejemplo, en [27], [23] y [10].

Definición 2.1.1. *Sean enteros m y n, con $2 \leq n \leq m$. Considere $\{v_1, v_2, \ldots, v_n\} \subseteq \mathbb{R}^m$ un conjunto de vectores linealmente independiente. El* ***lattice****, denotado por L o $L(v_1, v_2, \ldots, v_n)$, generado por $\{v_1, v_2, \ldots, v_n\}$ es el conjunto,*

$$L(v_1, v_2, \ldots, v_n) := \left\{ \sum_{i=1}^{n} m_i v_i : \ m_i \in \mathbb{Z} \right\},$$

de todas las combinaciones lineales enteras de $v_1, v_2, \ldots, v_n$.

Una ***base*** para L es cualquier conjunto de vectores $\{v_1, v_2, \ldots, v_n\}$ linealmente independiente que genera a L. Los enteros n y m son llamados ***rango y dimensión*** de L, respectivamente. Además si $n = m$, el lattice L es llamado de ***rango completo***.

Observación 2.1.1. *Sabemos que $\mathbb{R}^m$ es un espacio vectorial y en particular $(\mathbb{R}^m, +)$ es un grupo aditivo, con la suma usual de $\mathbb{R}^m$. Dado que $L(v_1, v_2, \ldots, v_n) \subset \mathbb{R}^m$, se puede probar que L es un subgrupo aditivo de $\mathbb{R}^m$, ya que L hereda las propiedades de grupo de $\mathbb{R}^m$.*

Consideremos la siguiente proposición.

Proposición 2.1.1. *Sean $\beta_1 = \{v_1, v_2, \ldots, v_n\}$ y $\beta_2 = \{w_1, w_2, \ldots, w_n\}$ dos bases de un lattice L. Entonces, β_1 y β_2 están relacionadas por una matriz A, que tiene coeficientes enteros y determinante igual a ± 1.*

Demostración. Sea $L \subset \mathbb{R}^n$ un lattice y supongamos que $\beta_1 = \{v_1, \ldots, v_n\}$ y $\beta_2 = \{w_1, \ldots, w_n\}$ son dos bases distintas para L. Considerando la matriz de cambio de base para β_1 y β_2 podemos escribir a los elementos, w_i, de β_2 como combinación lineal de los elementos de la base β_1, de la siguiente manera,

$$\begin{aligned} w_1 &= a_{11}v_1 + a_{12}v_2 + \cdots + a_{1n}v_n, \\ w_2 &= a_{21}v_1 + a_{22}v_2 + \cdots + a_{2n}v_n, \\ &\vdots \qquad\qquad\qquad \vdots \\ w_n &= a_{n1}v_1 + a_{n2}v_2 + \cdots + a_{nn}v_n \end{aligned}$$

El sistema de ecuaciones anterior se puede escribir matricialmente como $W = AV$, en donde los coeficientes a_{ij} son números enteros pues los w_k son elementos del lattice L.

Notemos que para poder expresar a los elementos v_i en términos de los w_k, se tiene que invertir la matriz $A = (a_{ij})$, para tener $V = A^{-1}W$. Esto quiere decir que necesitamos que los v_i sean combinaciones lineales de los w_k con coeficientes enteros, por lo que, las entradas de A^{-1} deben ser enteras. Es claro que A es invertible ya que es de rango completo, luego,

$$I_n = AA^{-1}, \text{ por lo que, } det(I_n) = det(AA^{-1}) = det(A)det(A^{-1}) = 1,$$

donde $det(A)$ y $det(A^{-1})$ son números enteros. Entonces, debemos tener que $det(A) = \pm 1$.

Por otro lado, uno de los conceptos estándares del álgebra lineal [33] dice que:

$$A^{-1} = \frac{Adj(A)}{det(A)}, \tag{2.1}$$

donde $Adj(A) = (ad_{ij})$ es la matriz adjunta de A y $det(A)$ es el determinante de la matriz A, con los coeficientes ad_{ij} dados por:

$$ad_{ij} = (-1)^{i+j} det(A_{ij}),$$

donde A_{ij} es la matriz que se obtiene al eliminar la fila i y la columna j de la matriz A.

Como $det(A) = \pm 1$ y si sustituimos este valor en la ecuación (2.1) obtenemos que,

$$A^{-1} = \pm 1 Adj(A),$$

con los coeficientes de A, a_{ij}, números enteros. Entonces claramente la matriz $Adj(A)$ forma una matriz con entradas de números enteros, esto a su vez nos dice que las coordenadas de la matriz A^{-1} son números enteros.

Por lo tanto, existe una matriz A con coeficientes enteros y determinante igual a ± 1 que relaciona a la base β_1 y β_2. □

De la proposición anterior establecemos la siguiente definición.

Definición 2.1.2. *Una matriz cuadrada A es **unimodular** si tiene todos sus coeficientes enteros y $det(A) = \pm 1$.*

Decimos que dos bases β_1 y β_2 para un lattice L son *equivalentes* si $L(\beta_1) = L(\beta_2)$. A continuación se ilustra la afirmación de la proposición 2.1.1 con el siguiente ejemplo.

Ejemplo 2.1.1. *Considere el lattice $L \subset \mathbb{R}^2$ generado por la base $\{v_1, v_2\}$, en donde $v_1 = (2,1), v_2 = (1,0)$.*

Formamos una matriz tomando los vectores $v_1 = (2,1)$ y $v_2 = (1,0)$, es decir,

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}.$$

Consideramos dos vectores en L mediante las fórmulas,

$$w_1 = v_1 = (2,1) \text{ y } w_2 = v_2 - v_1 = (-1,-1).$$

La matriz que forman los renglones w_1 y w_2 equivale a multiplicar la matriz A por una matriz U por la izquierda,

$$B = UA = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}.$$

La matriz U tiene determinante igual a 1, entonces los vectores w_1 y w_2 también son una base para L y la inversa de U es:

$$U^{-1} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Ahora, los renglones de U^{-1} nos dicen cómo expresar a los vectores v_1 y v_2 como combinación lineal de w_1 y w_2,

$$v_1 = 1w_1 + 0w_2 = (2,1) \quad y \quad v_2 = 1w_1 + 1w_2 = (1,0).$$

2.2. Interpretación geométrica de los determinantes

Sea $A = \begin{pmatrix} u_1 & u_2 \\ v_1 & v_2 \end{pmatrix}$. Los renglones, $u = (u_1, u_2)$ y $v = (v_1, v_2)$, de la matriz A determinan un paralelogramo, como se observa en la figura 2.1.

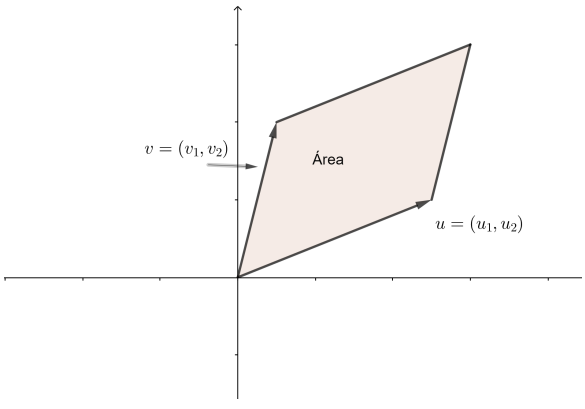


Figura 2.1: El área de la región sombreada es generada por los vectores u y v.

El área (volumen) generada por u y v se define como el área del paralelogramo dado en la figura 2.1. Podemos pensar que u y v son vectores en $\mathbb{R}^3$, que están en el plano xy. Entonces $u = (u_1, u_2, 0), v = (v_1, v_2, 0)$, y el área generada por u y v es la norma del producto,

$$|u \times v| = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & 0 \\ v_1 & v_2 & 0 \end{vmatrix} = |(u_1 v_2 - u_2 v_1)\mathbf{k}|.$$

Observe que $|det(A)| = |u \times v| = |u_1 v_2 - u_2 v_1|$. Por lo tanto, concluimos que el área generada por u y v es el valor absoluto del determinante de la matriz A.

Ahora si $\{u, v, w\}$ es una base de $\mathbb{R}^3$, entonces como en el caso de dimensión 2, estos vectores forman los lados de un paralelepípedo en el espacio $\mathbb{R}^3$. Calculemos su volumen.

La base del paralelepípedo es un paralelogramo de área igual a $|u \times v|$. El vector $u \times v$ es ortogonal tanto a u como a v, por lo tanto, es ortogonal al paralelogramo determinado por u y v.

La altura del paralelepípedo, h, se mide a lo largo del vector ortogonal al paralelogramo.

El escalar h, es el valor absoluto de la componente de w en la dirección (ortogonal) $u \times v$, es decir,

$$h = \text{componente de } w \text{ en la dirección } u \times v = \left| \frac{w \cdot (u \times v)}{|u \times v|} \right|.$$

Entonces,

$$\text{Volumen del paralelepípedo} = \text{área de la base} \times \text{altura} =$$

$$|u \times v| \left[\frac{|w \cdot (u \times v)|}{|u \times v|} \right] = |w \cdot (u \times v)|.$$

Es decir, el volumen del paralelepípedo determinado por los tres vectores u, v y w es igual a $|w \cdot (u \times v)|$, que es el valor absoluto del triple producto escalar de u, v y w.

Ampliamos el determinante a matrices de orden n generalizando la fórmula de volumen en $\mathbb{R}^2$ y $\mathbb{R}^3$ de la siguiente manera:

Sea $A = (a_{ij})$ una matriz cuadrada de orden n. Se define el *determinante de A como,*

$$det(A) = \sum_{\sigma \in S_n} signo(\sigma) \cdot a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdot \cdots \cdot a_{n,\sigma(n)},$$

en donde S_n es el grupo simétrico de orden n y $signo(\sigma)$ es el signo de la permutación σ.

2.3. Volumen de un lattice

Definición 2.3.1. *Sea L un lattice de dimensión n y sea $\{v_1, v_2, \ldots, v_n\}$ una base para L. El **dominio fundamental** (o paralelepípedo fundamental) para L, correspondiente a esta base es el conjunto,*

$$P(v_1, \ldots, v_n) = \{t_1v_1 + t_2v_2 + \cdots + t_nv_n : t_i \in \mathbb{R}, \text{ con } 0 \leq t_i < 1\}.$$

Proposición 2.3.1. *Sea $L \subset \mathbb{R}^n$ un lattice de dimensión n y sea F un dominio fundamental para L. Entonces, cualquier vector $w \in \mathbb{R}^n$ se puede escribir de manera única en la forma,*

$$w = t + v, \text{ para únicos } t \in F \text{ y } v \in L.$$

Demostración. Supongamos que los vectores $v_1, \ldots, v_n$ forman una base para el lattice L, con dominio fundamental F. Entonces $v_1, \ldots, v_n$ son linealmente independientes en $\mathbb{R}^n$, ya que son una base de $\mathbb{R}^n$. Luego, cualquier vector $w \in \mathbb{R}^n$ se puede escribir de la siguiente manera,

$$w = \alpha_1v_1 + \alpha_2v_2 + \cdots + \alpha_nv_n, \quad \text{para algunos } \alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}.$$

Notemos que cada α_i se puede escribir como,

$$\alpha_i = t_i + a_i, \quad \text{con } 0 \leq t_i < 1 \text{ y } a_i \in \mathbb{Z}.$$

Entonces,

$$w = \underbrace{t_1v_1 + t_2v_2 + \cdots + t_nv_n}_{\textit{vector en } F} + \underbrace{a_1v_1 + a_2v_2 + \cdots + a_nv_n}_{\textit{vector en } L}.$$

Esto prueba que w se puede escribir en la forma deseada.

Ahora, suponga que $w = t + v = t' + v'$ tiene dos representaciones como la suma de un vector en F y un vector en L. Entonces,

$$(t_1 + a_1)v_1 + (t_2 + a_2)v_2 + \cdots + (t_n + a_n)v_n = (t'_1 + a'_1)v_1 + (t'_2 + a'_2)v_2 + \cdots + (t'_n + a'_n)v_n.$$

Como los vectores $v_1, \ldots, v_n$ son linealmente independientes, tenemos que,

$$t_i + a_i = t'_i + a'_i, \text{ para toda } i = 1, 2, \ldots, n.$$

Notemos que los términos $t_i - t'_i = a'_i - a_i$ son enteros; además sabemos que t_i y t'_i son mayores o iguales que 0 y estrictamente menores que 1, por lo que, la única forma de que $t_i - t'_i$ sea un número entero es si $t_i = t'_i$. Por lo tanto, se tiene la igualdad $t = t'$ y también la igualdad en la expresión de v como suma de un elemento de F y un elemento de L, es decir,

$$v = w - t = w - t' = v'.$$

Esto prueba la unicidad. □

El *traslado de un dominio fundamental* F es una isometría T en el espacio euclidiano caracterizada por un vector u tal que para cada punto $p \in F$ le corresponde otro punto p' del espacio euclidiano, de modo que:

$$T_u : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$
$$p \mapsto p' = T(u) = p + u.$$

La proposición 2.3.1 equivale a considerar la unión de los dominios fundamentales trasladados,

$$v + F = \{t + v : t \in F\},$$

con v variando sobre los vectores del lattice L, es decir, $\mathbb{R}^n = \bigcup_{v \in L}(v + F)$.

Sea L un lattice de dimensión n y sea F un dominio fundamental para L. Entonces, ***el volumen del lattice*** L es igual al volumen del dominio fundamental F.

Dado que F es un paralelepípedo en $\mathbb{R}^n$, entonces de la definición 2.3.1 y por la sección 2.2, podemos escribir el volumen del lattice L en términos de un determinante, $vol(L) = vol(F) = det(A)$, donde A es la matriz que tiene por renglones a los vectores que forman una base para el lattice L.

Sabemos que dos bases equivalentes tienen el mismo determinante. Por lo tanto, el volumen de un paralelepípedo fundamental es independiente de la base elegida para el lattice L.

Cuando se tiene un lattice que no es de rango completo, es decir, el lattice tiene lugar en un espacio de dimensión menor que el espacio en donde pertenecen los vectores generadores del lattice L, se puede calcular el volumen del paralelepípedo fundamental, como se describe en [23] de la siguiente manera:

$$vol(L) = det(L(A)) = \sqrt{det(AA^T)}.$$

Definición 2.3.2. ***Los mínimos sucesivos asociados a un lattice*** *L de dimensión m y de rango n son las constantes $\lambda_1, \lambda_2, \ldots, \lambda_n$ definidas por,*

$$\lambda_i(L) = inf\,\{r : dim(gen(L \cap B_m(0, r))) \geq i\},$$

donde $B_m(0, r) = \{x \in \mathbb{R}^m : ||x|| < r\}$ es la bola abierta de radio r centrada en 0.

Es decir, el *i-ésimo* mínimo λ_i es el radio de la bola cerrada más pequeña centrada en el origen que contiene i vectores linealmente independientes del lattice L. Además es fácil ver que $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ y que λ_1 es la longitud del vector más corto distinto de cero en el lattice L.

En la teoría del álgebra lineal, un resultado muy útil y conocido es el proceso de ortogonalización de Gram–Schmidt, [33], el cual es un algoritmo para construir, a partir de un conjunto de vectores de un espacio vectorial con producto interno, otro conjunto ortonormal de vectores que genere el mismo subespacio vectorial. En el apéndice A se presenta el funcionamiento de este algoritmo.

Teorema 2.3.1. *Sea B una base para un lattice L y consideremos la base ortogonal B^* obtenida de la base B, mediante el proceso de Gram Schmidt. Entonces, el primer mínimo del lattice L con la norma euclidiana satisface que,*

$$0 < \min_j ||b_j^*|| \leq \lambda_1.$$

Demostración. Supongamos que el lattice L es de dimensión m y de rango n. Sea $x \in \mathbb{Z}^n$ un vector distinto de cero. Consideremos un vector arbitrario en el lattice L, digamos $b = xB$. Por definición de mínimos sucesivos asociados al lattice L, se tiene que $\lambda_1 = inf||b||$. Sea i el índice más grande tal que $x_i \neq 0$.

Notemos que $b = \sum_{j=1}^{i} b_j x_j$, luego,

$$\langle b, b_i^* \rangle = \sum_{j=1}^{i} \langle b_j, b_i^* \rangle x_j.$$

Dado que b_i^* es ortogonal a $b_1, b_2, \ldots, b_{i-1}$ y $\langle b_i, b_i^* \rangle = \langle b_i^*, b_i^* \rangle = ||b_i^*||^2$, tenemos que,

$$\langle b, b_i^* \rangle = \langle b_i, b_i^* \rangle x_i = ||b_i^*||^2 x_i.$$

Como $|x_i| \geq 1$, se sigue que $|\langle b, b_i^* \rangle| \geq ||b_i^*||^2$.

Por la desigualdad de Cauchy-Schwarz vemos que $|\langle b, b_i^* \rangle| \leq ||b||\ ||b_i^*||^2$, por eso $||b||\ ||b_i^*||^2 \geq ||b_i^*||^2$. Ahora, como cada vector b_i^* satisface que $||b_i^*|| \neq 0$, multiplicamos por $||b_i^*||^{-1}$ en está última desigualdad para obtener que $||b|| \geq ||b_i^*||$.

Luego, de la desigualdad $||b|| \geq ||b_i^*|| \geq \text{mín}_j\, ||b_j^*||$ tomamos el ínfimo de cada parte y así tenemos que,

$$\lambda_1 = inf||b|| \geq inf(\text{mín}_j\, ||b_j^*||) = \text{mín}_j\, ||b_j^*||.$$

□

2.4. El problema de vector más corto (SVP) y el problema de vector más cercano (CVP)

En esta sección presentamos dos de los problemas asociados a un lattice y que en particular son de los problemas computacionalmente difíciles de resolver. Estos conceptos son estándar y se pueden encontrar en [10].

Definición 2.4.1. ***El problema de vector más corto****, abreviado (SVP) por sus siglas en inglés, consiste en lo siguiente: Dado un lattice L, encontrar un vector distinto de cero $v \in L$ tal que $||v|| \leq ||w||$, para todo vector $w \in L$ distinto de cero.*

Definición 2.4.2. ***El problema de vector más cercano****, abreviado (CVP) por sus siglas en inglés, consiste en lo siguiente: Dado un lattice L y un vector $w \in \mathbb{R}^n$, encontrar un vector $v \in L$ que minimice la distancia $||w - v||$.*

Los dos problemas anteriores son muy difíciles de resolver, tanto que se han diseñado versiones aproximadas a estos problemas para reducir su dificultad. Presentamos dos versiones para estos problemas:

Definición 2.4.3. *(**Problema aproximado del problema SVP**) Dado un lattice L y un número real $\gamma > 1$, encontrar un vector distinto de cero $v \in L$ tal que $||v|| \leq \gamma||v'||$ para todo vector distinto de cero $v' \in L$.*

Definición 2.4.4. *(**Problema aproximado del problema CVP**) Dado un lattice L, un vector $t \in \mathbb{Z}^m$ y un número real $\gamma > 1$, encontrar un vector distinto de cero $v \in L$ tal que $||v - t|| \leq \gamma||v' - t||$ para todo vector distinto de cero $v' \in L$.*

Katherine Jarvis en su tesis, [20], da una explicación de cómo se puede resolver los problemas SVP y CVP para lattices rectangulares en $\mathbb{R}^2$, la cual consiste en lo siguiente:

Para resolver el problema CVP en lattices rectangulares en $\mathbb{R}^2$ consideramos un lattice rectangular $L \subset \mathbb{R}^2$ generado por una base ortogonal, digamos $B = \{(a, 0), (0, b)\}$, con $a, b \in \mathbb{R}$.Supongamos que deseamos resolver el problema CVP, para algún elemento $\alpha = (x, y) \in \mathbb{R}^2$ como se ilustra en la figura 2.2.

Entonces, consideramos el redondeo de coordenadas horizontales y verticales más cercanas al lattice L de la siguiente manera:

$$\gamma = (a\lfloor x/a \rceil, b\lfloor y/b \rceil) \quad \text{y} \tag{2.2}$$
$$\rho = (x - a\lfloor x/a \rceil, y - b\lfloor y/b \rceil), \tag{2.3}$$

donde $\lfloor \cdot \rceil$ representa la *función entero más cercano*, la cual se define por $\lfloor a \rceil = \lfloor a \rfloor$ si $|a - \lfloor a \rfloor| < 1/2$, y en caso contrario hacemos $\lfloor a \rceil = \lceil a \rceil$.

Resulta que el vector γ es un vector más cercano a $\alpha \in L$ y la expresión $||\rho|| = ||\alpha - \gamma||$ se minimiza.
Con lo anterior decimos que, de manera general, si β es una base para un lattice $L \subset \mathbb{R}^m$, podemos escribir en coordenadas relativas a una base ortonormal paralela a β y luego, usar una generalización de las coordenadas (2.2) y (2.4) para así encontrar el vector más cercano.

El problema SVP también se puede resolver de manera fácil en un lattice rectangular. Esto es, dada una base ortogonal $\beta = \{b_1, b_2, \ldots, b_n\}$ para un lattice $L \subset \mathbb{R}^n$, entonces cualquier vector distinto de cero x que está en L se puede escribir como combinación lineal de los vectores de β, es decir, $x = \sum_{i=1}^{n} a_i b_i$, con $a_i \in \mathbb{Z}$. Así que

$$||x||^2 = \sum_{i=1}^{n} |a_i|^2 ||b_i||^2 \geq \min_i \left\{ ||b_i||^2 \right\}.$$

Por lo tanto, se satisface el problema SVP.

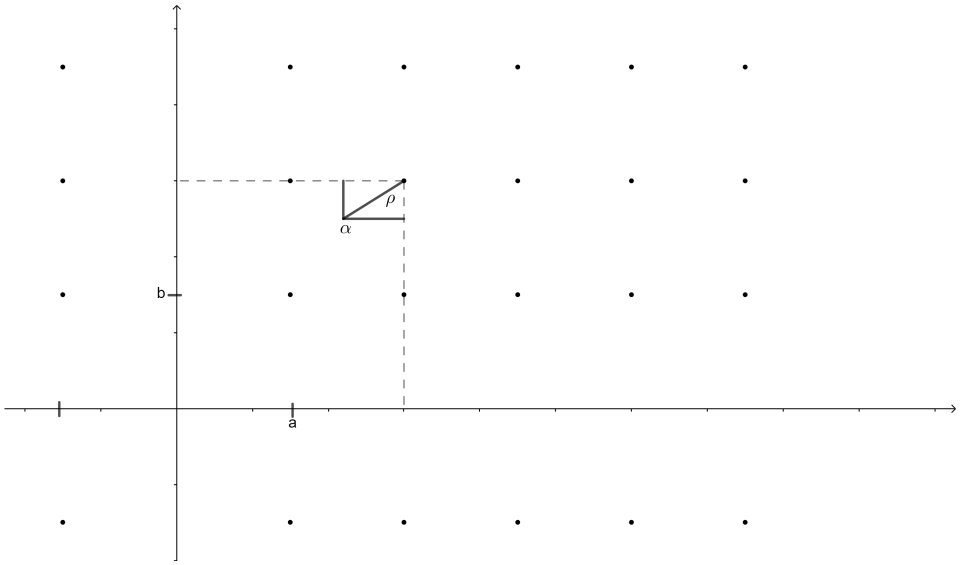


Figura 2.2: Gráfica del *lattice* L generado por la base $B = \{(a, 0), (0, b)\}$ y el vector α, el cual se considera como vector más cercano.

2.5. El algoritmo LLL

Dado que encontrar vectores cortos en un lattice es uno de los problemas más difíciles de resolver computacionalmente, ya que es un problema de tipo NP, algunos autores como C. Schnorr y M. Euchner, en [5], se han adentrado a buscar nuevos algoritmos que resuelvan el problema SVP de manera aproximada considerando la reducción de lattices. Un algoritmo eficiente que resuelve este tipo de problemas es el algoritmo LLL, y que además es de tiempo polinomial determinista, es decir, que tal algoritmo es factible o eficiente para algún ordenador.

Consideremos el espacio vectorial $\mathbb{R}^m$ con su producto interno usual, $\langle \cdot, \cdot \rangle$. Defina $\beta = \{b_1, b_2, \ldots, b_m\}$ una base de $\mathbb{R}^m$. Con el proceso de Gram Schmith, explicado en el apéndice A, podemos considerar la base ortogonal generada por este proceso, a partir de la base β, digamos $\beta^* = \{b_1^*, b_2^*, \ldots, b_m^*\}$. Con estos conceptos podemos dar la siguiente definición.

Definición 2.5.1. *Una base ordenada $\beta = \{b_1, b_2, \ldots, b_m\} \subseteq \mathbb{R}^m$ se dice que es de* ***tamaño reducido*** *si $|\mu_{i,j}| \leq 1/2$, para $1 \leq j < i \leq n$, donde $\mu_{i,j}$ son los coeficientes de Gram Schmidt.*

Definición 2.5.2. *Una base ordenada $\beta = \{b_1, b_2, \ldots, b_m\} \subseteq \mathbb{R}^m$ es llamada LLL-* ***reducida con parámetro*** *δ, donde $1/4 < \delta \leq 1$, si es de tamaño reducido y si para dos vectores consecutivos se cumple que,*

$$\delta ||b_i^*||^2 \leq ||b_{i+1}^* + \mu_{i+1,i} b_i^*||^2.$$

El siguiente lema presentado en [20] nos proporciona una cota superior para el primer vector de una base LLL-reducida.

Lema 2.5.1. *Sea $B = \{b_1, b_2, \ldots, b_m\} \subseteq \mathbb{R}^m$ una base LLL-reducida con parámetro $\frac{1}{4} < \delta \leq 1$. Entonces, $||b_1|| \leq \left(2/\sqrt{4\delta - 1}\right)^{n-1} \lambda_1$.*

Demostración. Por definición de base LLL-reducida tenemos que,

$$\delta||b_i||^2 \leq ||b_{i+1}^* + \mu_{i+1,i}b_i^*||^2, \text{ para todo } 1 \leq i < n.$$

Como b_i^* y b_{i+1}^* son ortogonales, se sigue que,

$$\delta||b_i||^2 = ||b_{i+1}^*||^2 + \mu_{i+1,i}^2||b_i^*||^2,$$

y dado que $|\mu_{i,j}| \leq 1/2$, para $j < i$; tenemos que $\delta||b_i||^2 \leq ||b_{i+1}^*|| + \frac{1}{4}||b_i^*||^2$, y reorganizando términos, se obtiene que $\left(\delta - \frac{1}{4}\right)||b_i^*||^2 \leq ||b_{i+1}^*||^2$.

Procediendo por inducción sobre el factor $i - j$ se probará que,

$$(\delta - 1/4)^{i-j}||b_j^*||^2 \leq ||b_i^*||^2, \quad \text{para toda} \quad i \geq j. \tag{2.4}$$

Para la base inductiva verificaremos que la afirmación se cumple para $i-j = 0$ y para $i - j = 1$.
Si $i - j = 0$, la desigualdad (2.4) se puede escribir de la siguiente manera:

$$(\delta - 1/4)^0||b_i^*||^2 \leq ||b_i^*||^2,$$

es decir, se tiene que $||b_i^*||^2 = ||b_i^*||^2$.

Si $i - j = 1$, tenemos lo siguiente,

$$\begin{aligned}(\delta - 1/4)^{i-(i-1)}||b_{i-1}^*||^2 &= (\delta - 1/4)||b_{i-1}^*||^2 \\ &= (\delta - 1/4)||b_j^*||^2 \\ &\leq ||b_{j+1}^*||^2 \\ &= ||b_i^*||^2.\end{aligned}$$

Por lo tanto, es válida la base inductiva.

En la hipótesis de inducción supongamos que la desigualdad (2.4) se cumple para toda $k \leq j \leq i$, es decir, se cumple que $(\delta - 1/4)^{i-j}||b_j^*||^2 \leq ||b_i^*||^2$. En el paso inductivo probaremos que también es válida para $j = k - 1$.

Vemos que $(\delta - 1/4)||b_{k-1}^*||^2 \leq ||b_k^*||^2$.
Si multiplicamos el factor $(\delta - 1/4)^{-i+j}$ de ambos lados en la hipótesis de inducción, obtenemos,

$$\begin{aligned}(\delta - 1/4)^{-i+j}(\delta - 1/4)^{i-j}||b_j^*||^2 &\leq (\delta - 1/4)^{-i+j}||b_i^*||^2 \\ \implies \quad ||b_j^*||^2 &\leq (\delta - 1/4)^{-i+j}||b_i^*||^2.\end{aligned}$$

En particular se cumple que $||b_k^*||^2 \leq (\delta - 1/4)^{-i+k}||b_i^*||^2$. Por lo tanto, concluimos que,

$$(\delta - 1/4)||b_{k-1}^*||^2 \leq ||b_k^*||^2 \leq (\delta - 1/4)^{-i+k}||b_i^*||^2.$$

De donde, resulta que,

$$(\delta - 1/4)^{i-(k-1)}||b_{k-1}^*||^2 \leq ||b_i^*||.$$

Por lo tanto, la desigualdad se cumple para toda $i \geq j$.

Dado que $b_1^* = b_1$ y por el teorema 2.3.1 ($\lambda_1 \geq min_i||b_i^*||^2$) se sigue que $(\delta - 1/4)^{n-1}||b_1^2||^2 \leq \lambda_1^2$, y entonces podemos concluir que,

$$||b_1|| \leq (\delta - 1/4)^{-(n-1)/2}\lambda_1 = \left(2/\sqrt{4\delta - 1}\right)^{n-1} \lambda_1.$$

□

Este resultado, como comenta K. Jarvis en su tesis, [20], proporciona un vector de un lattice, cuya norma es de alrededor $(2/\sqrt{4\delta - 1})^{n-1}$ del mínimo. Además, para $\delta = 0{,}99$ y considerando una base LLL-reducida nos dará un vector con un factor de aproximadamente $1{,}16^{n-1}$ del mínimo. Por lo tanto, es muy probable que una base LLL-reducida contenga vectores cortos para valores pequeños de n y valores grandes de δ.

El algoritmo LLL proporciona una base LLL-reducida de vectores relativamente cortos, a partir de una base $\beta = \{b_1, b_2, \ldots, b_n\}$ para un lattice L. Tal algoritmo consiste en los siguientes pasos:

1. Paso de reducción.
   Se reduce el tamaño de cada vector b_i, es decir, para $i = 1$ tomamos $b_1' = b_1$, luego, establecemos $i = 2$ y hacemos,

   $$b_i' = b_i - \sum_{j=1}^{i-1} \lfloor \mu_{i,j} \rceil b_j',$$

   donde $\lfloor \mu_{i,j} \rceil$ es la función entero más cercano a $\mu_{i,j}$, después, actualizamos los vectores b_i por b_i'.

2. Paso de intercambio.
   Si los vectores consecutivos b_i y b_{i-1} no satisfacen la condición de base LLL-reducida ($\delta||b_{i-1}^*||^2 \leq ||b_i^* + \mu_{i,i-1}b_{i-1}^*||^2$) cambiamos los vectores estableciendo el índice $i = max(i-1, 2)$ y actualizamos la base, y los coeficientes de Gram-Schmith correspondientes. De lo contrario aumente i en uno.

3. Repetir.
   Repita el paso 1. El algoritmo termina cuando $i = n$.

Una observación importante es que la reducción de tamaño de un vector b_k no afecta la reducción de tamaño de otros vectores, así que, cuando el algoritmo termina, tenemos una base de tamaño reducido y dado que la condición LLL es válida para toda i, en el paso 2, tenemos una base LLL-reducida. En general los algoritmos de reducción de base como el algoritmo LLL, entre otros, tienen buena oportunidad de encontrar un vector corto en un lattice.

La heurística gaussiana proporciona un valor aproximado de la longitud de un vector corto o (el más corto) en un lattice que se ha tomado al azar, [10]. Dicha longitud es,

$$s = \sqrt{\frac{d}{2\pi e}}(det(L))^{1/d},$$

donde d es la dimensión del lattice L. Este concepto será útil en las secciones 3.4, 6.6.2 y 7.3.3.

Capítulo 3

Criptosistema NTRU

En este capítulo presentamos el criptosistema NTRU con la selección de parámetros propuesta por Hoffstein, Pipher y Silverman en [10] y [13]. Dicha selección de parámetros asegura que el proceso no falle, sin embargo, cuando no se realiza así esta selección, el descifrado puede fallar con cierta probabilidad, la cual se calcula en la sección 3.4. También se estudia uno de los ataques comunes contra este cifrado, el ataque de lattices contra NTRU, que también se estudia en [20].

3.1. Criptosistema NTRU

3.1.1. Notación

Para desarrollar el mecanismo criptosistema NTRU, nombraremos a un emisor (persona que cifra algún mensaje) y a un receptor (persona que recibe y descifra el mensaje), por Aldo y Bianca, respectivamente.

El criptosistema NTRU es un sistema de cifrado de clave pública basado en polinomios, dicho cifrado ocupa tres anillos de polinomios, a saber, $R = \mathbb{Z}[x]/\langle x^n - 1\rangle$, $R_p = \mathbb{Z}_p[x]/\langle x^n - 1\rangle$ y $R_q = \mathbb{Z}_q[x]/\langle x^n - 1\rangle$; donde $\mathbb{Z}[x]$ es el anillo de polinomios con coeficientes enteros, $\langle x^n - 1\rangle$ es el *ideal generado por el polinomio* $x^n - 1$ y $\mathbb{Z}_p[x]$ es el anillo de polinomios con coeficientes en el anillo $\mathbb{Z}_p$.

Definición 3.1.1. *El* ***levantamiento de un polinomio*** $f(x) \in R_q$ *a* R *es el único polinomio* $f'(x) \in R$ *que satisface* $f'(x) \mod q = f(x)$*, cuyos coeficientes están en el intervalo* $\left(\frac{-q}{2}, \frac{q}{2}\right]$.

En otras palabras el levantamiento de un polinomio reduce los coeficientes del polinomio módulo q en el intervalo $\left(\frac{-q}{2}, \frac{q}{2}\right]$.

Definición 3.1.2. *Sean* d_1, d_2 *enteros positivos. Definimos el conjunto de* ***polinomios ternarios*** *como sigue:*

$$T(d_1, d_2) = \left\{ a(x) \in R : \begin{array}{l} a(x) \text{ tiene } d_1 \text{ coeficientes iguales a } 1 \\ a(x) \text{ tiene } d_2 \text{ coeficientes iguales a } -1 \\ a(x) \text{ tiene el resto de coeficientes iguales a } 0 \end{array} \right\}.$$

Entonces, considerando los anillos de polinomios R, R_p y R_q, elegimos cuatro parámetro n, p, q, d que satisfacen lo siguiente:

1. El parámetro n es primo.
2. El máximo común divisor de p y q es 1, al igual que el máximo común divisor de n y q, es decir, $(p, q) = (n, q) = 1$.
3. El parámetro d es un entero positivo.
4. El parámetro q debe satisfacer que $q > (6d + 1)p$.

Para facilitar los cálculos se sugiere tomar a p y q como números primos, pero esto no es indispensable en el cifrado.

3.1.2. Generación de claves

Para generar las claves del criptosistema NTRU, Aldo elige parámetros (n, p, q, d) que satisfagan las condiciones mencionadas anteriormente, luego, selecciona de manera aleatoria dos polinomios $f(x) \in T(d + 1, d)$ y $g(x) \in T(d, d)$. Aldo calcula (busca) los polinomios $F_p(x) = f^{-1}(x) \in R_p$ y $F_q(x) = f^{-1}(x) \in R_q$, es decir, los polinomio F_p y F_q deben satisfacer que,

$$F_p(x)f(x) \equiv 1 \text{ (mód } p) \text{ y } F_q(x)f(x) \equiv 1 \text{ (mód } q).$$

Observe que puede ocurrir que el polinomio $f(x)$ no tenga inverso en R_p o en R_q, en tal caso Aldo debe seleccionar otro polinomio $f(x)$ que sí tenga inverso multiplicativo en R_p y en R_q. Después, se calcula $h(x) = F_q(x)g(x) \mod q$. Note que la igualdad anterior equivale a tener $f(x)h(x) \equiv g(x) \text{ (mód } q)$.

Luego las claves de Aldo para NTRU son

1. Se deja como clave pública al polinomio $h(x)$.
2. Los parámetros n, p, q son públicos.
3. La clave privada de Aldo serán los polinomio $f(x)$ y $g(x)$.

En la práctica es importante tener almacenado, de manera segura, al polinomio $F_p(x)$, pues este es útil en el proceso de descifrado.

3.1.3. Cifrado

Supongamos que Bianca desea enviar un mensaje a Aldo. Primero debe buscar la clave pública de Aldo y después representa el mensaje como un elemento del anillo R, es decir, $m(x) \in R$, cuyos coeficientes están en el intervalo $\left(\frac{-p}{2}, \frac{p}{2}\right]$. Luego, elige un polinomio $r(x) \in R$ y calcula el siguiente polinomio,

$$e(x) = ph(x)r(x) + m(x) \mod q.$$

Así, el polinomio $e(x)$ es el mensaje cifrado.

Entonces, Bianca envía el mensaje cifrado $e(x)$ a Aldo.

3.1.4. Descifrado

Supongamos que Aldo recibe el mensaje de Bianca y para descifrarlo realiza los siguientes cálculos:

$$a(x) = f(x)e(x) \mod q,$$

luego, realiza el levantamiento del polinomio $a(x)$ a R obteniendo un polinomio $a'(x)$, y calcula el polinomio,

$$b(x) = F_p(x)a'(x) \mod p.$$

Se realiza el levantamiento de $b(x)$ a R, dando como resultado un polinomio $b'(x)$. De esta manera el mensaje descifrado es $m(x) = b'(x)$.

3.1.5. ¿Por qué funciona el descifrado?

Para verificar que el descifrado funciona, es decir, para ver si en verdad se ha recuperado el mensaje original $m(x)$ realizamos lo siguiente:

Dado que Aldo ha recibido el mensaje $e(x) = ph(x)r(x) + m(x)$ en R_q, sabe que $e(x)$ es un elemento de R_q, calcula lo siguiente,

$$\begin{aligned} a(x) &= f(x)e(x) \mod q \\ &\equiv f(x)[ph(x)r(x) + m(x)] \mod q \\ &= pf(x)h(x)r(x) + f(x)m(x) \mod q \\ &= pf(x)F_q(x)g(x)r(x) + f(x)m(x) \mod q \\ &= pg(x)r(x) + f(x)m(x) \mod q. \end{aligned}$$

Como los polinomios $g(x), r(x)$ están en $T(d, d)$, el mayor coeficiente en el polinomio $g(x)r(x)$ ocurre cuando en ambos polinomios el número de coeficientes 1 es igual al número de coeficientes -1, por lo que, el mayor coeficiente posible es $2d$ y de manera análoga $-2d$ es el menor coeficiente posible.

Dado que el polinomio $f(x) \in T(d+1, d)$ y los coeficientes de $m(x)$ están en el intervalo $\left(\frac{-p}{2}, \frac{p}{2}\right]$, entonces el mayor coeficiente que puede tener el polinomio $f(x)m(x)$ es $(2d+1)\frac{p}{2}$.

Por lo tanto, el mayor coeficiente posible en $a(x) = pg(x)r(x) + f(x)m(x) \mod q$ es,

$$p2d + (2d+1)\frac{p}{2} = p\left(3d + \frac{1}{2}\right) = \frac{1}{2}p(6d+1) < \frac{q}{2}.$$

Así, cuando Aldo calcula el polinomio $a(x)$ y realiza su levantamiento en R, obtiene un polinomio $a'(x) = pg(x)r(x) + f(x)m(x)$. Después de hacer el levantamiento, Aldo calcula lo siguiente,

$$\begin{aligned} b(x) &= F_p(x)a'(x) \mod p \\ &= F_p(x)[pg(x)r(x) + f(x)m(x)] \mod p \\ &= pF_p(x)g(x)r(x) + F_p(x)f(x)m(x) \mod p \\ &= pF_p(x)g(x)r(x) + m(x) \mod p \\ &= m(x) \mod p. \end{aligned}$$

Como los coeficientes de $m(x)$ están en el intervalo $\left(\frac{-p}{2}, \frac{p}{2}\right]$, al levantar el polinomio $b(x)$ a R se obtiene $b'(x)$, que resulta ser el mensaje $m(x)$. Por lo tanto, Aldo recupera el mensaje original $m(x)$.

3.2. Ejemplo del cifrado NTRU

Para ilustrar el funcionamiento del criptosistema NTRU, veamos el siguiente ejemplo.

Considere los siguientes parámetros,

$$\begin{aligned} n &= 7. \\ p &= 3. \\ q &= 79. \\ d &= 4. \\ f(x) &= -x^{12} - x^{11} + x^{10} - x^7 + x^6 + x^3 - x^2 + x + 1. \\ g(x) &= -x^{12} + x^{10} + x^9 + x^6 - x^5 + x^4 - x - 1. \\ m(x) &= -x^{12} - x^9 + x^7 + x^6 - x^5 - x^4 - x^2 - 1. \\ r(x) &= -x^{12} - x^{11} - 3x^{10} - x^9 - 2x^8 + 2x^6 - x^3 + x - 1. \end{aligned}$$

Con la elección de los parámetros anterior, obtenemos los siguientes polinomios del cifrado NTRU, con la ayuda del programa SageMath 9.0.

$$F_p(x) = 2x^{12} + 2x^{11} + x^{10} + x^9 + x^8 + x^7 + 2x^6 + x^5 + 2x^4 + x^3 + x + 1.$$

$$F_q(x) = 40x^{12} + 24x^{11} + 7x^{10} + 4x^9 + 56x^8 + 31x^7 + 17x^6 + 67x^5 + 12x^4 + 32x^3 + 8x^2 + 19.$$

$$h(x) = 78x^{12} + 18x^{11} + 25x^{10} + 40x^8 + 30x^7 + 2x^6 + 64x^5 + 58x^4 + 19x^3 + 65x^2 + 32x + 43.$$

Luego, el mensaje $m(x)$ queda cifrado de la siguiente manera,

$$e(x) = 38x^{12} + 36x^{11} + 49x^{10} + 23x^9 + 71x^8 + 16x^7 + 74x^6 + 43x^5 + 17x^4 + 77x^3 + 16x^2 + 31x + 58.$$

Ahora, realizamos el descifrado del mensaje $e(x)$. Primero calculamos el polinomio $a(x) = f(x)e(x) \mod q$, y se realiza el levantamiento de $a(x)$ a R,

$$a(x) = 6x^{12} + 15x^{11} + 20x^{10} + 8x^9 + 5x^8 + 71x^7 + 54x^6 + 61x^5 + 74x^4 + 2x^3 + 7x^2 + 66x + 2.$$

$$a'(x) = 6x^{12} + 15x^{11} + 20x^{10} + 8x^9 + 5x^8 - 8x^7 - 25x^6 - 18x^5 - 5x^4 + 2x^3 + 7x^2 - 13x + 2.$$

Finalmente calculamos el polinomio $b(x) = F_p(x)a'(x) \mod p$,

$$b(x) = 2x^{12} + 2x^9 + x^7 + x^6 + 2x^5 + 2x^4 + 2x^2 + 2,$$

después, realizamos el levantamiento de $b(x)$ a R y así obtenemos el mensaje original $m(x)$, es decir,

$$b(x) = -x^{12} - x^9 + x^7 + x^6 - x^5 - x^4 - x^2 - 1.$$

Calcular los tiempos de velocidad para cifrar y descifrar un mensaje con el criptosistema NTRU puede variar mucho por diferentes factores, como son el programa donde se implementó el criptosistema NTRU, la manera que se utilizó para multiplicar polinomios y aplicar reducciones modulares, el equipo de cómputo del que se dispone, etc. K. Jarvis en su tesis [20] realiza una estimación de tiempo que toma cifrar y descifrar 10000 mensajes, haciendo notar que cifrar resulta relativamente más rápido que descifrar un mensaje.

3.3. Fallo de descifrado

Para realizar el proceso de descifrado de NTRU, Aldo tiene que calcular el polinomio $A = a(x) \equiv f(x)e(x) \equiv pg(x)r(x) + f(x)m(x) \mod q$. Si Aldo eligió los parámetros (n, p, q, d) como lo indica la sección 3.1.2, el descifrado nunca falla.

Tomemos en cuenta que en la práctica puede haber varios usuarios que utilicen el criptosistema NTRU para intercambiar información, por lo que si otro usuario, digamos Bianca, desea enviar algún mensaje a estos usuarios, entonces Bianca tiene que realizar tablas que expresen la codificación del mensaje deseado con sus respectivos parámetros elegidos por un primer usuario y así enviar su mensaje cifrado $e_1(x)$ a este usuario.

Si un primer usuario ha elegido distintos parámetros que un segundo usuario, Bianca se ve en la necesidad de volver a calcular nuevas tablas (evaluaciones de polinomios) para poder cifrar su mensaje y que el descifrado sea correcto. por lo tanto, si no se realiza una revisión de parámetros antes de cifrar un mensaje, se tiene el riesgo de enviar mensajes cifrados con distintos parámetros a distintos usuarios. Este escenario puede causar un error en el proceso de descifrado, cuando un mensaje cifrado es recibido por los usuarios en cuestión.

También, si algún usuario llegase a tomar algún parámetro que no satisfaga las condiciones de la sección 3.1.2, el descifrado puede no ser el correcto. Por estas razones y posiblemente otros argumentos de cuestiones prácticas, se puede tener la probabilidad de que los coeficientes del polinomio A no se encuentren en el intervalo $\left(-\frac{q}{2}, \frac{q}{2}\right]$, cuando se aplica la reducción módulo q. Entonces, no se recupera el mensaje original $m(x)$ y en este caso tendríamos que $A \neq a(x) \equiv f(x)e(x) \equiv pg(x)r(x) + f(x)m(x) \mod q$. Veamos cuál es la probabilidad de que falle el descifrado.

La forma explícita del polinomio A es:

$$\begin{aligned} A &= pg(x)r(x) + f(x)m(x) \mod q \\ &= p(g_0 + g_1x + g_2x^2 + \cdots + g_{n-1}x^{n-1})(r_0 + r_1x + r_2x^2 + \cdots + r_{n-1}x^{n-1}) + \\ &\quad (f_0 + f_1x + f_2x^2 + \cdots + f_{n-1}x^{n-1})(m_0 + m_1x + m_2x^2 + \cdots + m_{n-1}x^{n-1}) \mod q \\ &= p\sum_{j=0}^{n-1}\sum_{k+\ell\equiv j}(g_kx^k)(r_\ell x^\ell) + \sum_{j=0}^{n-1}\sum_{k+\ell\equiv j}(f_kx^k)(m_\ell x^\ell) \\ &= p\sum_{j=0}^{n-1}\sum_{k+\ell\equiv j} g_kr_\ell x^j + \sum_{j=0}^{n-1}\sum_{k+\ell\equiv j} f_km_\ell x^k, \end{aligned}$$

en donde $k+\ell \equiv j$ representa la congruencia $k+\ell \equiv j \pmod n$. Si a los términos u_kx^k de cada polinomio los representamos simplemente por u_k, con esta notación tenemos que,

$$A = p\sum_{j=0}^{n-1}\sum_{k+\ell\equiv j} g_kr_\ell + \sum_{j=0}^{n-1}\sum_{k+\ell\equiv j} f_km_\ell.$$

El j-ésimo término del polinomio A es,

$$A_j = p \sum_{k+\ell \equiv j} g_k r_\ell + \sum_{k+\ell \equiv j} f_k m_\ell.$$

Sin pérdida de generalidad podemos suponer que los A_j son variables aleatorias independientes que se distribuyen de manera normal.

Sabemos que las unidades del anillo $\mathbb{Z}$ son $U(\mathbb{Z}) = \{1, -1\}$ y que los polinomios $r(x)$ y $g(x)$ están en $T(d_r, d_r)$ y $T(d_g, d_g)$, respectivamente. Esto quiere decir que los coeficientes de $r(x)$ y $g(x)$ están en el conjunto $\{1, -1, 0\}$.

La probabilidad de que ocurran los eventos $g_k = b$ y $r_\ell = b$, donde $b \in \{1, -1, 0\}$ es $P(g_k = b) = \frac{d_g}{n}$ y $P(r_\ell = b) = \frac{d_r}{n}$, respectivamente.

Calculamos las siguientes probabilidades,

$$\begin{aligned} P(g_k r_\ell = 1) &= P(g_k r_\ell = -1) \\ &= \sum_{r \in U(\mathbb{Z})} P(r_\ell = r, g_k = r^{-1}) \\ &= \sum_{r \in U(\mathbb{Z})} P(r_\ell) P(g_k = r^{-1}) \\ &= 2\left(\frac{d_g}{n}\right)\left(\frac{d_r}{n}\right) = \frac{2d_g d_r}{n^2}. \end{aligned}$$

Entonces, el valor esperado y varianza del término $g_k r_\ell$ son,

$$E(g_k r_\ell) = 1\left(\frac{2d_g d_\ell}{n^2}\right) + (-1)\left(\frac{2d_g d_\ell}{n^2}\right) = 0.$$

$$\begin{aligned} Var(g_k r_\ell) &= E((g_k r_\ell)^2) - E(g_k r_\ell)^2 \\ &= E((g_k r_\ell)^2) + 0 \\ &= (1)^2\left(\frac{2d_g d_\ell}{n^2}\right) + (-1)^2\left(\frac{2d_g d_\ell}{n^2}\right) \\ &= \frac{4d_g d_r}{n^2}. \end{aligned}$$

La otra componente del polinomio A es el producto de los polinomios $m(x)$ y $f(x)$, donde $m(x) \in T(d_m, d_m)$ y $f(x) \in T(d_f + 1, d_f)$. Consideremos los siguientes eventos: $f_k = -1$ y $f_k = 1$. Entonces, sus respectivas probabilidades son: $P(f_k = -1) = \frac{d_f}{n}$ y $P(f_k = 1) = \frac{d_f+1}{n}$, que para un valor de n suficientemente grande la aproximación $P(f_k = 1) = \frac{d_f+1}{n} \approx \frac{d_f}{n}$ es aceptable.

Dado que en este caso quisiéramos que el polinomio $m(x)$ tenga sus coeficientes enteros en el intervalo $\left(-\frac{q}{2}, \frac{q}{2}\right]$, entonces podemos considerar el evento $m_\ell = \lambda$, donde λ puede

tomar cualquier valor en el conjunto $\left\{-\frac{q-1}{2}, -\frac{q-1}{2}+1, \ldots, \frac{q-1}{2}\right\}$ y así tenemos que la probabilidad de que suceda el evento m_ℓ es $P(m_\ell = \lambda) = \frac{1}{q}$. De modo que,

$$\begin{aligned} P(f_k m_\ell = \lambda) &= \sum_{u \in U(\mathbb{Z})} P(f_k = u, m_\ell = u^{-1}\lambda) \\ &= \sum_{u \in U(\mathbb{Z})} P(f_k = u) P(m_\ell = u^{-1}\lambda) \\ &= 2\left(\frac{d_f}{n}\right)\left(\frac{1}{q}\right) = \frac{2d_f}{nq}. \end{aligned}$$

Calculamos el valor esperado y la varianza del término $f_k m_\ell$,

$$E(f_k m_\ell) = (-1)\sum_{i=1}^{\frac{q-1}{2}} i\left(\frac{2d_f}{nq}\right) + (1)\sum_{i=1}^{\frac{q-1}{2}} i\left(\frac{2d_f}{nq}\right) = 0.$$

$$\begin{aligned} Var(f_k m_\ell) &= E((f_k m_\ell)^2) - E(f_k m_\ell)^2 = E((f_k m_\ell)^2) - 0 \\ &= \left(-\frac{q-1}{2}\right)^2\left(\frac{2d_f}{nq}\right) + \left(-\frac{q-1}{2}+1\right)^2\left(\frac{2d_f}{nq}\right) + \cdots + \left(\frac{q-1}{2}\right)^2\left(\frac{2d_f}{nq}\right) \\ &= 2\sum_{i=1}^{\frac{q-1}{2}} i^2\left(\frac{2d_f}{nq}\right) \\ &= 2\left(\frac{2d_f}{nq}\right)\frac{\frac{q-1}{2}\left(\frac{q-1}{2}+1\right)\left(2\frac{(q-1)}{2}+1\right)}{6} \\ &= 2\left(\frac{2d_f}{nq}\right)\frac{\frac{q-1}{2}(\frac{q+1}{2})(q)}{6} \\ &= \frac{4d_f}{nq}\left[\frac{(q-1)(q+1)(q)}{4}/6\right] \\ &= \frac{d_f(q-1)(q+1)}{6n}. \end{aligned}$$

Con los resultados calculados anteriormente y aplicando la linealidad del operador esperanza o valor esperado, procedemos a calcular el valor esperado y la varianza del j-ésimo coeficiente A_j.

$$\begin{aligned} E(A_j) &= E\left(p\sum_{k+\ell \equiv j} g_k r_\ell + \sum_{k+\ell \equiv j} f_k m_\ell\right) \\ &= p\sum_{k+\ell \equiv j} E(g_k r_\ell) + \sum_{k+\ell \equiv j} E(f_k m_\ell) \\ &= p\sum_{k+\ell \equiv j} 0 + \sum_{k+\ell \equiv j} 0 = 0. \end{aligned}$$

Con el supuesto de que los términos $g_k r_\ell$ y $f_k m_\ell$ son variables aleatorias independientes, aplicamos las propiedades básicas de varianza, vistas en [26] para obtener,

$$\begin{aligned} Var(A_j) &= Var\left(p\sum_{k+\ell\equiv j} g_k r_\ell + \sum_{k+\ell\equiv j} f_k m_\ell\right) \\ &= p^2 \sum_{k+\ell\equiv j} Var(g_k r_\ell) + \sum_{k+\ell\equiv j} Var(f_k m_\ell) \\ &= p^2 n\frac{4d_g d_r}{n^2} + n\frac{d_f(q-1)(q+1)}{6n} \\ &= \frac{4p^2 d_g d_r}{n} + \frac{d_f(q-1)(q+1)}{6} \\ &= \frac{4p^2 g_k r_\ell}{n} + \frac{d_f(q^2-1)}{6}. \end{aligned}$$

Es común denotar al valor esperado y varianza de una variable aleatoria como μ y σ^2 respectivamente. Esta notación es estándar y se puede consultar en las referencias [31] y [26]. Además, como estamos suponiendo que los coeficientes de A_j se distribuyen de manera normal, entonces la probabilidad de que el coeficiente A_j se encuentre en el intervalo $\left(-\frac{q}{2}, \frac{q}{2}\right]$ es,

$$\begin{aligned} P\left(|A_j| \leq \frac{q-1}{2}\right) &= P\left(-\frac{q-1}{2} \leq A_j \leq \frac{q-1}{2}\right) \\ &= P\left(\frac{-\frac{q-1}{2}-\mu}{\sigma} \leq \frac{A_j-\mu}{\sigma} \leq \frac{\frac{q-1}{2}-\mu}{\sigma}\right) \\ &= P\left(\frac{-q+1-2\mu}{2\sigma} \leq Z \leq \frac{q-1-2\mu}{2\sigma}\right) \\ &= P\left(\frac{-q+1}{2\sigma} \leq Z \leq \frac{q-1}{2\sigma}\right) \\ &= \Phi\left(\frac{q-1}{2\sigma}\right) - \Phi\left(-\frac{q-1}{2\sigma}\right) \\ &= 2\Phi\left(\frac{q-1}{2\sigma}\right) - 1, \end{aligned}$$

donde $\sigma = \sqrt{Var(A_j)}$ es la desviación estándar de A_j, Z es una variable aleatoria normal estándar, es decir $Z \sim N(0,1)$ y $\Phi(\cdot)$ es la función de distribución normal estándar.

Entonces, con lo anterior podemos decir que la probabilidad de que todos los coeficientes del polinomio A se encuentren en el intervalo $\left[-\frac{q}{2}, \frac{q}{2}\right]$ es,

$$P\left(||A||_\infty \leq \frac{q-1}{2}\right) = \left(2\Phi\left(\frac{q-1}{2\sigma}\right) - 1\right)^n,$$

donde $||A||_\infty = \max\{|A_0|, |A_1|, \ldots, |A_{n-1}|\}$. Observe que a medida que q aumenta, la probabilidad de que el descifrado falle disminuye y a medida que d_f, d_g, d_r incrementan, la probabilidad de descifrado también disminuye.

3.4. Análisis de seguridad

La funcionalidad del criptosistema NTRU está basada en anillos de polinomios y la relación entre la clave pública y privada se puede utilizar para construir un lattice, denominado el lattice estándar NTRU y denotado por L^{NT}. Una base para este lattice puede derivarse de la clave pública, por lo que, queda disponible públicamente.

La clave privada del cifrado NTRU la podemos considerar como un vector corto en este lattice. Se han propuesto varios ataques para encontrar la clave privada de NTRU, por ejemplo, el ataque de fuerza bruta, el ataque de encuentro en el medio, etc. Estos ataques, entre otros, se estudian en [20]. En esta sección veremos uno de los principales ataques al criptosistema NTRU, llamado Ataque de lattices, que también es estudiado en la referencia ya mencionada.

Por definición, la clave pública de NTRU está dada por $h(x) = F_q(x)g(x) \mod q$ o equivalentemente $f(x)h(x) = g(x) \mod q$. Dado un polinomio $f(x)$, denotamos su vector correspondiente de coeficientes como f. Esto es, si $f(x) = f_0 + f_1x + \cdots + f_nx^n$, entonces $f = (f_0, f_1, \ldots, f_n)$.

Consideramos el lattice L^{NT} de la siguiente manera:

$$L^{NT} = \{(u, v) | u(x), v(x) \in R_q \text{ y } u(x)h(x) \equiv v(x) \pmod q\}.$$

Entonces, es claro que $L^{NT} \subseteq \mathbb{Z}^{2n}$ es un lattice y además el vector (f, g) está en el lattice L^{NT}.

Para encontrar una base de L^{NT} observamos que $f(x)h(x) \equiv g(x) \pmod q$ se puede escribir como $f(x)h(x) - u(x)q = g(x)$, para algún polinomio $u(x) \in R$. Así que, en términos de coeficientes de los polinomios anteriores, tenemos la siguiente expresión en forma matricial,

$$[f, -u] \begin{bmatrix} \lambda I & H \\ 0 & qI \end{bmatrix} = [\lambda f, fh - qu] = [\lambda f, g] \in L^{NT}, \tag{3.1}$$

en donde λ es una constante real distinta de cero, I es una matriz identidad de tamaño $n \times n$ y H es la matriz circulante de tamaño $n \times n$ con primer renglón los coeficientes de $h(x)$.

La expresión anterior generaliza el lattice L^{NT}, pues para cada valor fijo de λ se tendrá un lattice que amplifique o comprima al lattice L^{NT}. Cuando se tiene el caso en que $\lambda = 1$ es por qué la elección de parámetros para el cifrado NTRU se ha realizado como en la sección 3.1.2, y más específicamente los polinomios $f(x)$ y $g(x)$ se consideran elementos de polinomios ternarios $T(d, d)$, siempre que n sea suficientemente grande.

Con un cálculo rutinario de bases para un espacio vectorial, es fácil probar que las filas de la matriz,

$$B^{NT} = \begin{bmatrix} \lambda I & H \\ 0 & qI \end{bmatrix},$$

forman una base para el lattice L^{NT}. De este modo vemos que $u^T H = uh$, para cualquier polinomio $u(x) \in R$.

Por lo tanto, el lattice estándar NTRU, L^{NT}, es un lattice de dimensión $2n$ que es generado por una matriz de tamaño $2n \times 2n$, denotada por B^{NT} y que de manera explícita tiene la forma,

$$B^{NT} = \left[\begin{array}{cccc|cccc} \lambda & 0 & \cdots & 0 & h_0 & h_1 & \cdots & h_{n-1} \\ 0 & \lambda & \cdots & 0 & h_{n-1} & h_0 & \cdots & h_{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda & h_1 & h_2 & \cdots & h_0 \\ \hline 0 & 0 & \cdots & 0 & q & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & q & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & q \end{array}\right],$$

donde los h_i son los coeficientes de $h(x)$, la clave pública NTRU, y $\lambda \in \mathbb{R}$ es una constante fija. La elección de la constante λ se puede realizar de manera conveniente para poder encontrar la clave privada NTRU.

De esta manera si logramos recuperar el vector $[f, g]$, entonces podemos recuperar la clave privada de NTRU, $f(x), g(x)$. De forma similar a lo anterior es fácil deducir que todos los vectores en el lattice L^{NT} pueden escribirse de la forma $[\lambda f', g']$, donde $g'(x) \equiv f'(x)h(x) \pmod{q}$.

Dado que los polinomios $f(x)$ y $g(x)$ son elementos de conjuntos de polinomios ternarios, claramente todos sus coeficientes tienen norma 0 o 1, entonces el vector $[f, g]$ estará entre los vectores cortos en el lattice L^{NT}. Por lo tanto, podemos aplicar técnicas de reducción de lattices, como el algoritmo LLL, para recuperar el vector $[\lambda f, g]$ u otros vectores cortos de la forma $[\lambda f', g']$, como las rotaciones de f y g.

Veamos la siguiente proposición propuesta por K. Jarvis, que identifica vectores cortos en el lattice L^{NT}.

Proposición 3.4.1. *Suponga que $f(x), g(x)$ son polinomios ternarios. Entonces, el vector $(\underbrace{\lambda, \ldots, \lambda}_{n \text{ veces}}, \underbrace{0, \ldots, 0}_{n \text{ veces}})$ está contenido en el lattice NTRU, L^{NT}.*

Demostración. Dado que B^{NT} es una base para el lattice L^{NT}, cualquier vector en L^{NT} se puede expresar como combinación lineal de los renglones de B^{NT}.

Si consideramos la suma de los primeros n renglones de B^{NT}, obtenemos el vector,

$$(\underbrace{\lambda, \ldots, \lambda}_{n \text{ veces}}, \underbrace{k, \ldots, k}_{n \text{ veces}}),$$

donde k representa la suma de los coeficientes en el polinomio $h(x)$. Esto significa que solo debemos probar que la suma de los coeficientes de $h(x)$ es 0 módulo q.

Como $h(x) \equiv F_q(x)g(x) \pmod{q}$, se sigue que $h(x) = F_q(x)g(x) + u(x)(x^n - 1)$, para algún $u(x) \in R_q$. Observemos que $g(x)$ tiene $d_g 1's$ y $d_g - 1's$, por lo que, $g(1) = 0$. Por lo tanto, $h(1) \equiv F_q(1)g(1) + u(1)(0) \equiv 0 \pmod{q}$. Esto quiere decir que los coeficientes de $h(x)$ suman 0 mód q. □

Por lo tanto, el vector $\underbrace{(\lambda, \ldots, \lambda}_{n \text{ veces}}, \underbrace{0, \ldots, 0)}_{n \text{ veces}}$ es otro vector en el lattice L^{NT}. Además, este vector puede ser más corto que nuestro vector objetivo $(\lambda f, g)$.

3.4.1. Elección de la constante de equilibrio λ

La constante λ es una constante de equilibrio que se utiliza para maximizar la eficiencia de la búsqueda de vectores cortos en el lattice L^{NT}.

Las técnicas de reducción de lattices, por ejemplo, la implementación del algoritmo LLL, tendrá la mayor oportunidad de encontrar el vector de longitud objetivo τ, o un vector cuya longitud sea cercana a τ, si τ es mucho más corto que el vector esperado en L^{NT}.

Por lo tanto, se desea elegir λ que maximice la relación s/τ, donde s es la longitud del vector esperado más corto en L^{NT}. En la sección 2.5 del capítulo 2 se expresa que el vector más corto distinto de cero en un lattice L dado por la heurística gaussiana es de longitud,

$$s = \sqrt{\frac{d}{2\pi e}}(det(L))^{1/d},$$

donde d es la dimensión del lattice L.

Entonces, para el caso del lattice L^{NT} tenemos que la dimensión de L^{NT} es $d = 2n$, $det(B^{NT}) = \lambda^n q^n$, de ahí que la longitud esperada del vector distinto de cero más corto en L^{NT} es:

$$s = \sqrt{\frac{d}{2\pi e}}(det(L))^{1/d} = \sqrt{\frac{2n}{2\pi e}}(\lambda^n q^n)^{1/2n} = \sqrt{\frac{n\lambda q}{\pi e}}.$$

La longitud de nuestro vector objetivo $(\lambda f, g)$ es:

$$\tau = \sqrt{\lambda^2 ||f(x)||^2 + ||g(x)||^2}.$$

Si deseamos maximizar la relación,

$$s/\tau = \sqrt{\frac{n\lambda q}{\pi e}} / \sqrt{\lambda^2 ||f(x)||^2 + ||g(x)||^2} = \sqrt{\frac{nq}{\pi e}} \sqrt{\frac{\lambda}{\lambda^2 ||f(x)||^2 + ||g(x)||^2}},$$

entonces, maximizar el término anterior es lo mismo que maximizar,

$$\lambda/(\lambda^2 ||f(x)||^2 + ||g(x)||^2) = (\lambda ||f(x)||^2 + \lambda^{-1} ||g(x)||^2)^{-1}.$$

Un cálculo fácil nos permite ver que la derivada con respecto a λ del término a maximizar, $\lambda/(\lambda^2 ||f(x)||^2 + ||g(x)||^2)$, es,

$$-(\lambda ||f(x)||^2 + \lambda^{-1} ||g(x)||^2)^{-2}(||f(x)||^2 - \lambda^{-2} ||g(x)||^2).$$

Por lo tanto, para maximizar la relación deseada se debe cumplir que,

$$||f(x)||^2 - \lambda^{-2} ||g(x)||^2 = 0 \quad \text{o bien} \quad \lambda = ||g(x)||/||f(x)||,$$

el cual es un máximo. Notemos que si $d_f = d_g$, es decir, si consideramos que los polinomios $f(x), g(x)$ están en $T(d, d)$, tendremos que $\lambda = 1$. En la práctica si $d_f = d_g$, entonces $||f(x)|| \approx ||g(x)||$ y se toma $\lambda = 1$.

Cualquier método para resolver el problema SVP puede romper el cifrado NTRU, es decir, encontrar la clave privada $(f(x), g(x))$. Actualmente no existe una forma eficiente capaz de resolver el problema SVP. Alternativamente se cuenta con el algoritmo LLL, que puede ser útil para resolver el problema aproximado SVP. Se han realizado experimentos en los que se muestran que para un valor de n pequeño, el algoritmo LLL es suficiente para romper el cifrado NTRU, [12], pero esto no representa un problema grave para valores de n que se utilizan en la práctica, los cuales son valores relativamente grandes.

Capítulo 4

Propiedades de los Enteros de Eisenstein

En este capítulo se presentan algunas de las propiedades básicas del anillo de los enteros de Eisenstein, las cuales serán de gran utilidad para implementar el criptosistema NTRU sobre los enteros de Eisenstein, este material está basado en [17], [4] y [20].

4.1. Los enteros de Eisenstein $\mathbb{Z}[\omega]$

Dentro de la teoría de números existen estructuras algebraicas llamadas anillos de enteros, que son de gran importancia en esta teoría. Por ejemplo, un anillo de enteros es $\mathbb{Z}$, dotado de varias propiedades:

1. El algoritmo de la división depende esencialmente de la función valor absoluto $|\cdot| : \mathbb{Z} \longrightarrow \mathbb{N}$.

2. 1 y -1 dividen a cualquier entero y son los únicos enteros que tienen inverso multiplicativo.

3. Si p es un número primo, entonces formalmente $-p$ también es un número primo.

4. $\mathbb{Z}$ es de factorización única.

En este capítulo estudiaremos el anillo de los enteros de Eisenstein, el cual surge de la siguiente manera:

Un número algebraico es cualquier número real o complejo que es solución de una ecuación algebraica de la forma:

$$z^m + \alpha_{m-1}z^{m-1} + \cdots + \alpha_1 z + \alpha_0 = 0,$$

donde $m > 0$ es el grado del polinomio, $\alpha_i \in \mathbb{Q}$. El conjunto de los números algebraicos es un campo.

Ahora bien, un número algebraico a, se dice que es un *entero algebraico* si satisface una ecuación de la forma $z^m + \alpha_{m-1}z^{m-1} + \cdots + \alpha_1 z + \alpha_0 = 0$, donde $\alpha_0, \ldots, \alpha_{m-1}$ son enteros.

El concepto de número algebraico es una motivación para hablar de las soluciones de la ecuación $x^4 - 1 = 0$. Esta ecuación tiene soluciones sobre los números complejos a los números ± 1, $\pm i$.

Consideremos las raíces cúbicas de la unidad, que son soluciones de la ecuación $x^3 = 1$. Dado que el polinomio $x^3 - 1 = (x-1)(x^2+x+1)$ tiene como raíces a 1 y $(-1 \pm i\sqrt{3})/2$. Si definimos $\omega = (-1 + i\sqrt{3})/2$, entonces es fácil comprobar que ω satisface la ecuación $1 + x + x^2 = 0$.

El conjunto $\mathbb{Z}[\omega] = \{a + b\omega : a, b \in \mathbb{Z}\}$ es cerrado bajo suma y diferencias de números complejos. Además, el producto de dos números en $\mathbb{Z}[\omega]$ se define de forma usual como sigue,

$$(a + b\omega)(c + d\omega) = ac + (ad + bc)\omega + bd\omega^2 = (ac - bd) + (ad + bc - bd)\omega.$$

Por lo tanto, $\mathbb{Z}[\omega]$ es un anillo y dado que $\mathbb{Z}[\omega]$ es un subconjunto de los números complejos, es un dominio entero.

Observamos que $\mathbb{Z}[\omega]$ es cerrado bajo conjugación compleja pues $\bar{\omega} = (-1 - i\sqrt{3})/2$, y del hecho de que $\omega^2 + \omega + 1 = 0$, tenemos que $\omega^2 = -1 - \omega = (-1 - i\sqrt{3})/2$, es decir, se cumple que $\bar{\omega} = \omega^2$. En consecuencia si $z = a + b\omega \in \mathbb{Z}[\omega]$, entonces

$$\bar{z} = a + b\bar{\omega} = a + b\omega^2 = (a - b) - b\omega \in \mathbb{Z}[\omega].$$

4.1.1. Divisibilidad en $\mathbb{Z}[\omega]$

Una vez que conocemos las operaciones de suma y producto en $\mathbb{Z}[\omega]$, en particular el producto, podemos definir el concepto de divisibilidad para lo cual necesitamos un algoritmo de la división. En el siguiente resultado se define una función, llamada *función norma*, que dota al anillo de los enteros de Eisenstein ser un dominio euclidiano.

Proposición 4.1.1. *El anillo de los enteros de Eisenstein $\mathbb{Z}[\omega]$ es un anillo euclidiano.*

Demostración. Para $\alpha = a + b\omega \in \mathbb{Z}[\omega]$, tenemos que,

$$\begin{aligned}
\alpha\bar{\alpha} &= (a + b\omega)[(a - b) - b\omega] \\
&= a^2 - ab - ab\omega + ab\omega - b^2\omega - b^2\omega^2 \\
&= a^2 - ab - b^2\omega - (b^2(-1 - \omega)) \\
&= a^2 - ab - b^2\omega + b^2 + b^2\omega \\
&= a^2 - ab + b^2.
\end{aligned}$$

Entonces, definimos $\nu(\alpha) = a^2 - ab + b^2 = \alpha\bar{\alpha}$.

Fijamos los elementos $\alpha, \beta \in \mathbb{Z}[\omega]$, con $\beta \neq 0$. Entonces, $\alpha/\beta = \alpha\bar{\beta}/(\beta\bar{\beta}) = c + d\omega$, donde $c, d \in \mathbb{Q}$. En lo anterior, hemos utilizado el hecho de que $(\beta\bar{\beta}) = \nu(\beta)$ es un entero positivo y que $\alpha\bar{\beta} \in \mathbb{Z}[\omega]$, ya que $\mathbb{Z}[\omega]$ es cerrado bajo productos.

Ahora, buscamos enteros m, n tales que $\mid c-m \mid \leq \frac{1}{2}$ y $\mid d-n \mid \leq \frac{1}{2}$. Defina $z = m+n\omega$, entonces,

$$\begin{aligned}\nu((\frac{\alpha}{\beta} - z)) &= (c-m)^2 - (c-m)(d-n) + (d-n)^2 \\ &\leq \frac{1}{4} \pm \frac{1}{4} + \frac{1}{4} < 1.\end{aligned}$$

Si definimos $\rho = \alpha - z\beta \in \mathbb{Z}[\omega]$, se sigue que $\rho = 0$ o bien

$$\nu(\rho) = \nu(\beta(\tfrac{\alpha}{\beta} - z)) = \nu(\beta)\nu(\tfrac{\alpha}{\beta} - z) < \nu(\beta).$$

Por lo tanto, la función ν convierte a $\mathbb{Z}[\omega]$ en un anillo euclidiano. □

Recordemos que el conjunto de números complejos $\mathbb{C}$ también tiene la estructura de espacio vectorial sobre $\mathbb{R}$ y por tanto podemos hablar de bases para subconjuntos de $\mathbb{C}$, por ejemplo, una base para los enteros de Eisenstein es $B = \{1, \omega\} \subset \mathbb{Z}[\omega] \subset \mathbb{C}$ y dado que $\mathbb{Z}[\omega]$ es cerrado bajo sumas, el conjunto de todas las combinaciones lineales enteras de 1 y ω están en $\mathbb{Z}[\omega]$. Por lo tanto, el conjunto $\mathbb{Z}[\omega]$ es un lattice en $\mathbb{C}$ generado por la base $B = \{1, \omega\}$.

Dado que $\mathbb{C} \cong \mathbb{R}^2$, es decir, $\mathbb{C}$ es isomorfo a $\mathbb{R}^2$ como espacios vectoriales, los elementos $\{1, \omega\}$ en la base $\{1, i\}$ se representan en $\mathbb{R}^2$ con los vectores $(1, 0)$ y $(-1/2, \sqrt{3}/2)$, además, es fácil verificar que el ángulo que separa a estos vectores es de 120°, y tienen la misma longitud. La base $\{1, \omega\}$ forma un lattice hexagonal como se muestra en la siguiente figura.

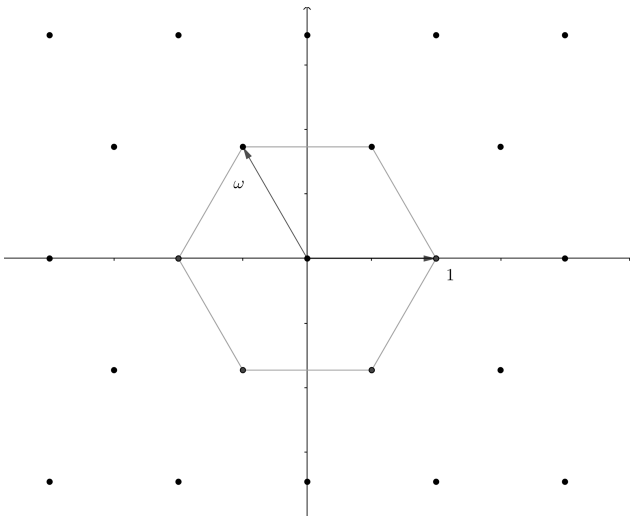


Figura 4.1: Los puntos en el plano complejo representan los enteros de Eisenstein, y el hexágono representa una celda del lattice generado por la base $\{1, \omega\}$.

Si un número eiseniano β es distinto de cero, *el ideal generado por* β es el conjunto,

$$\begin{aligned}(\beta) &= \{\gamma\beta | \gamma \in \mathbb{Z}[\omega]\} \\ &= \{(a + b\omega)\beta | a, b \in \mathbb{Z}\} \\ &= \{a\beta + b\beta\omega | a, b \in \mathbb{Z}\} \\ &= L(\beta, \beta\omega).\end{aligned}$$

Claramente (β) es un lattice en $\mathbb{Z}[\omega]$ generado por la base $\{\beta, \beta\omega\}$.

Definición 4.1.1. *Un dominio fundamental, H, del lattice (β) es un subconjunto de $(\beta) \subset \mathbb{Z}[\omega]$, que contiene un único representante de cada una de las distintas clases de congruencia módulo b.*

Por ejemplo, se puede verificar que si $\beta = 1 + 3\omega$, entonces una región fundamental H del ideal (β) consta de 0 y las unidades de $\mathbb{Z}[\omega]$.

La siguiente figura ilustra una región fundamental para el ideal (q), con $q = 4 + 5\omega$.

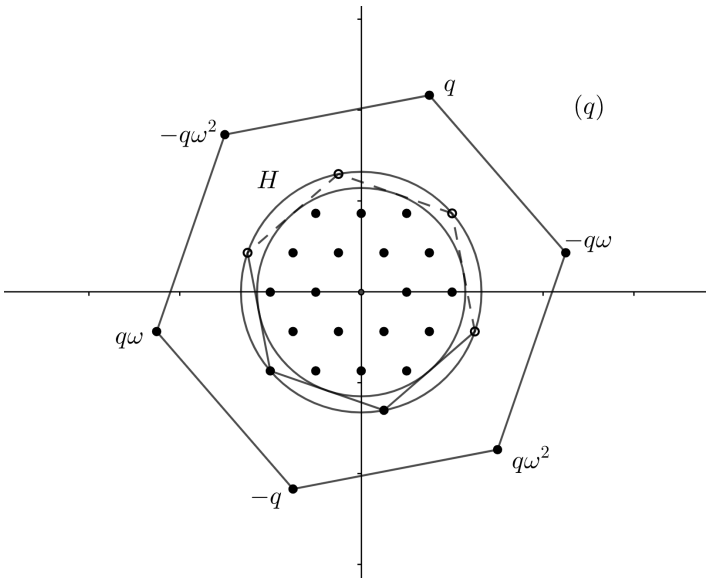


Figura 4.2: Los vértices $\pm q, \pm q\omega, \pm q\omega^2$, que forman un hexágono, son elementos del lattice (q), con $q = 4+5\omega$. Los puntos o vértices, que forman el hexágono H, junto con los puntos en el interior de H, representa un dominio fundamental de (q), el cual se ubica entre los círculos de radio $r_1 = \sqrt{\nu(q)}/2$ y $r_2 = \sqrt{\nu(q)/3}$.

Presentamos un algoritmo de la división en el anillo $\mathbb{Z}[\omega]$, desarrollado por K. Jarvis en [20], que se basa en su estructura que tiene como lattice. Esto es, dados $\alpha, \beta \in \mathbb{Z}[\omega]$, con β distinto de cero, queremos encontrar elementos $\gamma, \rho \in \mathbb{Z}[\omega]$, tales que $\alpha = \beta\gamma + \rho$.

Sabemos que la dimensión del espacio vectorial $\mathbb{C}$ sobre el campo $\mathbb{R}$ es dos. Podemos considerar la transformación lineal $T : \mathbb{C} \longrightarrow \mathbb{C}$ dada por $T(z) = \beta z$, donde $\beta = c + di$ es fijo. Entonces, aplicando la evaluación de T a los elementos de la base $\mathcal{B} = \{1, i\}$ de $\mathbb{C}$, obtenemos que,

$$T(1) = \beta = c + di,$$
$$T(i) = \beta i = -d + ci.$$

Por lo tanto, la representación matricial de T en la base $\mathcal{B}$ es,

$$[T]_{\mathcal{B}} = \begin{bmatrix} c & -d \\ d & c \end{bmatrix}.$$

Así, para $z = x + yi \in \mathbb{C}$, lo podemos identificar con el vector $z = \binom{x}{y}^T \in \mathbb{R}^2$. Se sigue que $z\beta = z[T]_{\mathcal{B}}$.

Consideramos la función parte entera más cercana, $\lfloor a \rceil$, en donde como antes si se cumple que $|a - \lfloor a \rfloor| < 1/2, \lfloor a \rceil = \lfloor a \rfloor$, de lo contrario hacemos $\lfloor a \rceil = \lceil a \rceil$. También, usaremos la notación $\langle \beta \rangle_{\mathcal{B}}$ o simplemente $\langle \beta \rangle$ para referirnos a la matriz $[T]_{\mathcal{B}}$, generada por el número complejo $\beta = c + di$ en la base $\{1, i\}$. Más adelante, en la sección 5.3 del capítulo 5, usaremos esta misma notación para definir el producto matricial que representa al producto de los números eisenianos y en tal caso se especificará el producto con su respectiva base $\{1, \omega\}$.

Teorema 4.1.1. *(Algoritmo de la división) Sea β y α números eisenianos, con $\beta \neq 0$. Definimos a_1, a_2, b_1, b_2 números reales tales que:*

$$a_1 + b_1 i = \beta^{-1}\alpha \qquad \text{y} \qquad a_2 + b_2 i = \beta^{-1}\alpha - \omega.$$

Para $j = 1, 2$; calcule,

$$\rho'_j = (a_j - \lfloor a_j \rceil) + i(b_j - \sqrt{3}\lfloor b_j/\sqrt{3} \rceil).$$

Defina $\rho_1, \rho_2, \gamma_1, \gamma_2$ números eisenianos como sigue,

$$\begin{aligned} \rho_1 &= \beta\rho'_1, \qquad \gamma_1 = \lfloor a_1 \rceil + i\sqrt{3}\lfloor b_i/\sqrt{3} \rceil \quad \text{y} \\ \rho_2 &= \beta\rho'_2, \qquad \gamma_2 = \lfloor a_2 \rceil + i\sqrt{3}\lfloor b_2/\sqrt{3} \rceil + \omega. \end{aligned}$$

Entonces, se cumple lo siguiente,

$$\alpha = \beta\gamma_1 + \rho_1 = \beta\gamma_2 + \rho_2 \quad \text{y} \quad Re(\gamma_1) \neq Re(\gamma_2),$$

donde $Re(\gamma)$ denota la parte real de $\gamma \in \mathbb{Z}[\omega]$.

Luego, defina (ρ, γ) como sigue:

- *Si $\nu(\rho_1) < \nu(\rho_2)$, seleccione $(\rho, \gamma) = (\rho_1, \gamma_1)$.*
- *Si $\nu(\rho_1) > \nu(\rho_2)$, seleccione $(\rho, \gamma) = (\rho_2, \gamma_2)$.*
- *Si $\nu(\rho_1) = \nu(\rho_2)$ y $Re(\gamma_1) > Re(\gamma_2)$, seleccione $(\rho, \gamma) = (\rho_1, \gamma_1)$.*
- *Si $\nu(\rho_1) = \nu(\rho_2)$ y $Re(\gamma_1) < Re(\gamma_2)$, seleccione $(\rho, \gamma) = (\rho_2.\gamma_2)$,*

El proceso que se genera cuando se aplica al par (α, β) es un algoritmo de la división, y se mantienen las siguientes condiciones:

1. *$\alpha = \beta\gamma + \rho \quad y \quad \nu(\rho) < \nu(\beta)$.*
2. *$\gamma\beta$ es un elemento del lattice $L(\beta)$ más cercano a α.*
3. *ρ es un representante más pequeño de la clase de congruencia módulo β.*
4. *El algoritmo de la división genera un representante único de la clase de congruencia, es decir si $\alpha, \alpha' \in \mathbb{Z}[\omega]$ satisfacen que $\alpha \equiv \alpha'$ (mód β), entonces el algoritmo produce el mismo resultado cuando se aplica a (α, β) y (α', β).*

Demostración. Primero se verificará que $\alpha = \beta\gamma + \rho$ en cualquiera de los dos casos propuestos en este algoritmo, de la siguiente manera:

Para el caso donde $\rho = \rho_1 = \beta\rho_1'$ y $\gamma = \gamma_1$ se tiene que,

$$\begin{aligned}\gamma\beta + \rho &= \beta(\lfloor a_1 \rceil + i\sqrt{3}\lfloor b_1/\sqrt{3} \rceil) + \beta((a_1 - \lfloor a_1 \rceil) + i(b_1 - \sqrt{3}\lfloor b_1/\sqrt{3} \rceil)) \\ &= \beta(a_1 + b_1 i) \\ &= \alpha, \qquad \text{ya que } (a_1 + b_2 i) = \beta^{-1}\alpha.\end{aligned}$$

Similarmente se puede verificar el caso en el que se establece $\rho = \rho_2 = \beta\rho_2'$ y $\gamma = \gamma_2$.

Ahora, verificaremos que $\gamma\beta$ es un elemento de $L(\beta)$ más cercano a α.

En la sección 2.4 del capítulo 2 vimos que el punto del lattice más cercano a un lattice rectangular se puede encontrar fácilmente con coordenadas relativas a la base ortogonal. Podemos transformar el lattice generado por la base $\{\beta, \beta\omega\}$ en el lattice generado por la base $\{1, \omega\}$, multiplicando por β^{-1}. Entonces, podemos encontrar el vector más cercano a $\beta^{-1}\alpha$ en el lattice $\{1, \omega\}$, que es el conjunto de los enteros de Eisenstein, de la siguiente manera:

Notemos que los enteros de Eisenstein se pueden ver como la unión de un lattice rectangular L_R, generado por la base $\left\{1, \sqrt{3}\right\}$ y su clase lateral $L_R + \omega$. Ahora bien, el punto más cercano en un lattice rectangular se da redondeando cada una de las coordenadas al número entero más cercano, de modo que ρ_1 determina el vector al punto del lattice más cercano en L_R a $a_1 + b_1 i$.

Análogamente el vector ρ_2' determina el vector a un punto del lattice más cercano γ_2' en L_R a $v_2 = a_2 + b_2 i$, que es equivalente a encontrar el vector al punto de lattice más cercano en $L_R + \omega$ (por la manera en que se definió el elemento γ_2) a $v_1 = a_1 + b_1 i$. Por lo tanto, $\gamma\beta$ es el punto del lattice más cercano en (β) a α.

De lo anterior se deduce que el número real $\nu(\rho)$ es el elemento mínimo del conjunto $\{\nu(x) | x \in \alpha + L(\beta)\}$ o equivalentemente ρ es un representante más pequeño de la clase de congruencia módulo β. También es claro que $\nu(\rho) < \nu(\beta)$.

Ahora veamos que el algoritmo de la división siempre genera un único representante de la clase de congruencia. Verificaremos que si $\alpha, \alpha' \in \mathbb{Z}[\omega]$ satisfacen que $\alpha \equiv \alpha'$ (mód β), entonces el algoritmo produce el mismo resultado cuando se aplica a (α, β) y (α', β).

Sabemos que este algoritmo de la división, 4.1.1, genera vectores más cercanos, así que solo se necesita considerar los casos en los que es equidistante a dos o más puntos en $L(\beta)$, cuando aplicamos dicho algoritmo a un par (α, β); de manera equivalente, asumimos que $\beta^{-1}\alpha$ es equidistante de dos o más números enteros de Eisenstein.

En el caso que haya dos puntos más cercanos en L_R equidistantes de $\beta^{-1}\alpha$, la consistencia de la función parte entera más cercano (redondeado a 0,5 hacia arriba) en ρ_1, garantiza que siempre elijamos el punto más cercano en L_R que este más a la derecha (o hacia arriba) de $\beta^1\alpha$ y similarmente ρ_2 siempre elije el punto más cercano a la derecha

(o hacia arriba) de $\beta^{-1}\alpha$ en $L_R + \omega$.

En el caso $\nu(\rho_1) = \nu(\rho_2)$, los elementos más cercanos $\beta^{-1}\alpha$ en L_R y $L_R + \omega$ son equidistantes de $\beta^{-1}\alpha$ y siempre elegimos el representante más a al derecha, es decir, si $Re(\gamma_1) > Re(\gamma_2)$, entonces seleccionamos $(\rho, \gamma) = (\rho_1, \gamma_1)$, de lo contrario si $Re(\gamma_1) < Re(\gamma_2)$, seleccionamos $(\rho, \gamma) = (\rho_2, \gamma_2)$.

Como L_R y $L_R+\omega$ son conjuntos disjuntos, $Re(\gamma_1) \neq Re(\gamma_2)$. Por lo tanto, el algoritmo de la división siempre genera un representante único, ya que siempre elegimos el punto más cercano a $\beta^{-1}\alpha$ más a la derecha (o hacia arriba). □

Veamos un ejemplo del funcionamiento del algoritmo de la división visto en el teorema anterior,

Ejemplo 4.1.1. *Dados $\alpha = 9 + 8\omega$ y $\beta = 1 + 3\omega$ en $\mathbb{Z}[\omega]$, calculemos α* mód *β.*

De acuerdo al algoritmo descrito en el resultado anterior, se debe calcular sucesivamente,

$$\alpha\beta^{-1} = a_1 + ib_1 = \tfrac{31}{14} - i\tfrac{19\sqrt{3}}{14}.$$
$$\alpha\beta^{-1} - \omega = a_2 + ib_2 = \tfrac{19}{7} - i\tfrac{13\sqrt{3}}{7}.$$
$$\rho_1' = \tfrac{31}{14} - \left\lfloor \tfrac{31}{14} \right\rceil + i\left(\tfrac{19\sqrt{3}}{14} - \sqrt{3}\left\lfloor \tfrac{19}{14} \right\rceil\right) = \tfrac{3}{14} - i\tfrac{5\sqrt{3}}{14}.$$
$$\rho_2' = \tfrac{19}{7} - \left\lfloor \tfrac{19}{7} \right\rceil + i\left(\tfrac{13\sqrt{3}}{7} - \sqrt{3}\left\lfloor \tfrac{13}{7} \right\rceil\right) = -\tfrac{2}{7} + i\tfrac{\sqrt{3}}{7}.$$
$$\rho_1 = \beta\left(\tfrac{3}{14} - i\tfrac{5\sqrt{3}}{14}\right) = \tfrac{3}{2} - i\tfrac{\sqrt{3}}{2}.$$
$$\gamma_1 = \left\lfloor \tfrac{31}{14} \right\rceil + i\sqrt{3}\left\lfloor \tfrac{19}{14} \right\rceil = 2 - i\sqrt{3}.$$
$$\rho_2 = \beta\left(-\tfrac{2}{7} + i\tfrac{\sqrt{3}}{7}\right) = -\tfrac{1}{2} - i\tfrac{\sqrt{3}}{2}.$$
$$\gamma_2 = \left\lfloor \tfrac{19}{7} \right\rceil + i\sqrt{3}\left\lfloor \tfrac{13}{7} \right\rceil + \omega = \tfrac{5}{2} - i\tfrac{3\sqrt{3}}{2}.$$

Como

$$\nu(\rho_1) = 3 > \nu(\rho_2) = 1,$$

se toma $(\rho, \gamma) = (\rho_2, \gamma_2)$. Con estos cálculos se concluye que,

$$\alpha \text{ mód } \beta = \rho = -\frac{1}{2} - i\frac{\sqrt{3}}{2} = -1 - \omega = \omega^2.$$

Ahora que se tiene un algoritmo de la división en $\mathbb{Z}[\omega]$ es posible utilizarlo para dar una definición de divisibilidad en este anillo. Sean $z, w \in \mathbb{Z}[\omega]$ con $w \neq 0$ y tal que $z = wk + \delta$, donde k y δ son números eisenianos. Decimos que w divide a z si $\delta = 0$. En tal caso si w divide a z, escribiremos $w|z$, y $w \nmid z$ en caso contrario.

Teorema 4.1.2. *Sean $z_1, z_2, z_3 \in \mathbb{Z}[\omega]$. Se cumplen las siguientes afirmaciones:*

1. *Si $z_1 \neq 0$, entonces $z_1|0, 1|z_1, z_1|z_1$.*
2. *Si $z_1|z_2$ y $z_2|z_3$, entonces $z_1|z_3$.*

3. *Si* $z|x_1, z|x_2, \ldots, z|x_n$, *entonces* $z|\sum_{i=1}^{n} a_i x_i$, *para todo* $a_i \in \mathbb{Z}[\omega]$.

4. *Si* $z_1|z_2$, *entonces* $\bar{z_1}|\bar{z_2}$.

5. *Si* $z_1|z_2$, *entonces* $\nu(z_1)|\nu(z_2)$.

Demostración. Para la propiedad 1 tenemos que dado $z_1 = a + b\omega$ y el elemento cero de $\mathbb{Z}[\omega]$ denotado por $0 = 0 + 0\omega$, entonces se tiene que,

$$z_1 0 = (a + b\omega)(0 + 0\omega) = (a0 - b0) + (a0 + b0 - b0)\omega = 0 + 0\omega = 0.$$

Se sigue que $z_1|0$.

Para la propiedad 2 definimos $z_1 = a + b\omega$, $z_2 = c + d\omega$, $z_3 = g + h\omega$. Como $z_1|z_2$ y $z_2|z_3$, entonces $z_2 = z_1 u_1$ y $z_3 = z_2 u_2$ para algunos $u_1, u_2 \in \mathbb{Z}[\omega]$, luego,

$$z_3 = z_2 u_2 = (z_1 u_1) u_2 = z_1 (u_1 u_2),$$

de donde $z_1|z_3$.

Para la propiedad 3, tenemos que $z|x_1, z|x_2, \ldots, z|x_n$, entonces,

$$x_1 = zu_1,\ x_2 = zu_2, \ldots,\ x_n = zu_n,$$

para algunos $u_1, u_2, \ldots, u_n \in \mathbb{Z}[\omega]$. Sean $v_1, v_2, \ldots, v_n \in \mathbb{Z}[\omega]$, luego tenemos que,

$$x_1 v_1 = z u_1 v_1,\ x_2 v_2 = z u_2 v_2, \ldots,\ x_n v_n = z u_n v_n.$$

Es decir, $\sum_{i=1}^{n} x_i v_i = z(\sum_{i=1}^{n} u_i v_i)$, lo que implica que $z|\sum_{i=1}^{n} x_i v_i = \sum_{i=1}^{n} v_i x_i$.

Para la propiedad 4, tenemos que si $z_1 = a + b\omega$, $z_2 = c + d\omega$ y $z_3 = a' + b'\omega$ es tal que $z_2 = (a + b\omega)(a' + b'\omega) = (aa' - bb') + (ab' + ba' - bb')\omega$, entonces,

$$\begin{aligned}
\bar{z_2} &= [aa' - bb' - (ab' + ba' - bb')] - (ab' + ba' - bb')\omega \\
&= [aa' - bb' - ab' - ba' + bb'] - (ab' + ba' - bb')\omega \\
&= (-ab' - ba' + aa') - (ab' + ba' - bb')\omega \\
&= (aa' - ba' - ab' + bb' - bb') + (-ab' + bb' - ba' + bb' - b'b)\omega \\
&= [(a - b)(a' - b') - (-b)(-b)'] + [(a - b)(-b)' + (-b)(a' - b') - (-b)(-b)']\omega \\
&= [(a - b) - b\omega][(a' - b') - b'\omega] \\
&= \bar{z_1}\bar{z_3}.
\end{aligned}$$

Se sigue que $\bar{z_1}|\bar{z_2}$.

Supongamos que $z_1|z_2$, entonces $z_2 = z_1 u_1$, para algún $u_1 \in \mathbb{Z}[\omega]$. Aplicamos la función ν en ambos lados de la igualdad anterior y obtenemos,

$$\nu(z_2) = \nu(z_1 u_1) = \nu(z_1)\nu(u_1),$$

lo que implica que $\nu(z_1)|\nu(z_2)$. □

En la prueba de la propiedad 4 del teorema anterior se verifico que $\bar{z_1}|\bar{z_2}$, es decir, $\bar{z_2} = \bar{z_1}z$, con $z \in \mathbb{Z}[\omega]$. Esto quiere decir que z es el conjugado del entero de Eisenstein z_3 en la hipótesis $z_2 = z_1z_3$. Además, el conjunto $\mathbb{C}$ satisface la distribución bajo conjugación, es decir, que el conjugado de un producto es el producto de los conjugados; claramente esta propiedad se hereda a $\mathbb{Z}[\omega]$.

El siguiente resultado nos hace notar que la divisibilidad en $\mathbb{Z}$ es consecuencia de la definición de divisibilidad en $\mathbb{Z}[\omega]$.

Teorema 4.1.3. *Sean $a, b \in \mathbb{Z}$ tal que $a|b$ en $\mathbb{Z}[\omega]$, con $a \neq 0$. Entonces $a|b$ en $\mathbb{Z}$.*

Demostración. Por hipótesis $b = az$ en $\mathbb{Z}[\omega]$ y $z \in \mathbb{Z}[\omega]$. Digamos $z = c + d\omega$, luego

$$\begin{aligned} b + 0\omega &= (a + 0\omega)(c + d\omega) \\ &= (ac - 0d) + (ad + 0c - 0d)\omega \\ &= ac + ad\omega. \end{aligned}$$

Igualando partes imaginarias tenemos que $0 = ad$, y como $a \neq 0$, entonces se sigue que $d = 0$. Así que $b = ac + ad\omega = ac$. Por lo tanto, $a|b$ en $\mathbb{Z}$. □

Algunos enteros dividen a cualquier elemento de $\mathbb{Z}[\omega]$, por ejemplo, la enunciado 1 del teorema 4.1.2 afirma que el número 1 divide a cualquier entero de Eisenstein. En el anillo de enteros de Eisenstein diremos que un elemento $z_1 \in \mathbb{Z}[\omega]$ es una *unidad* si $z_1|z_2$, para todo $z_2 \in \mathbb{Z}[\omega]$. Podemos caracterizar las unidades con la ayuda de la función norma.

Teorema 4.1.4. *Sea $z \in \mathbb{Z}[\omega]$. Entonces z es unidad si y solo si $\nu(z) = 1$.*

Demostración. Si z es unidad, en particular $z|1$ y por tanto $1 = zu$, para algún $u \in \mathbb{Z}[\omega]$. De lo anterior se sigue que,

$$1 = \nu(1) = \nu(zu) = \nu(z)\nu(u).$$

Así que $\nu(z)|1$, y como $\nu(z)$ es un entero positivo, entonces $\nu(z) = 1$. Recíprocamente si $\nu(z) = z\overline{z} = 1$, entonces para todo $z_2 \in \mathbb{Z}[\omega]$ se tiene que $z(\overline{z}z_2) = z_2$, es decir, $z|z_2$. □

Corolario 4.1.1. *Las unidades de $\mathbb{Z}[\omega]$ son $1, -1, \omega, -\omega, \omega^2, -\omega^2$.*

Demostración. Supongamos que $z = a + b\omega$ es una unidad. Entonces $1 = a^2 - ab + b^2$ o bien $4 = (2a - b)^2 + 3b^2$, luego, las soluciones en $\mathbb{Z}$ de la ecuación anterior son:

1. $(a, b) = (\pm 1, 0)$.
2. $(a, b) = (0, \pm 1)$.
3. $(a, b) = (\pm 1, \pm 1)$.

De lo anterior se obtiene que $1, -1, \omega, -\omega, \omega^2, -\omega^2$ son soluciones en $\mathbb{Z}[\omega]$. □

Una observación importante es que precisamente las unidades de $\mathbb{Z}[\omega]$ son los únicos elementos de $\mathbb{Z}[\omega]$ que tienen inverso multiplicativo. El conjunto de las unidades de $\mathbb{Z}[\omega]$ lo denotaremos por $U(\mathbb{Z}[\omega])$. Es fácil verificar que el conjunto $U(\mathbb{Z}[\omega])$ forma un grupo bajo el producto de $\mathbb{Z}[\omega]$.

Sean $z_1, z_2 \in \mathbb{Z}[\omega]$. Diremos que z_1 y z_2 son *asociados* si $z_1|z_2$ y $z_2|z_1$. Notemos que si z_1 y z_2 son asociados, entonces $z_1 = z_2z_3$, donde z_3 es alguna unidad y que por el corolario 4.1.1, cada entero eiseniano distinto de 0 tiene exactamente seis asociados.

Ejemplo 4.1.2. *Los números $z_1 = 2 - 5\omega$ y $z_2 = 5 + 7\omega$ son asociados, pues se cumple que,*

$$\begin{aligned}(5+7\omega)(\omega^2) &= 5\omega^2 + 7\omega^3 \\ &= 5(-1-\omega) + 7 \\ &= -5 - 5\omega + 7 \\ &= 2 - 5\omega \in \mathbb{Z}[\omega],\end{aligned}$$

y también,

$$\begin{aligned}(2-5\omega)(\omega) &= 2\omega - 5\omega^2 \\ &= 2\omega - 5(-1-\omega) \\ &= 2\omega + 5 + 5\omega \\ &= 5 + 7\omega \in \mathbb{Z}[\omega].\end{aligned}$$

Ejemplo 4.1.3. *Si $z = 5+2\omega$, entonces los asociados de z son: $z, -z, \omega z, -\omega z, \omega^2 z, -\omega^2 z$.*

Al conjunto $\mathbb{Z}[\omega]$ es imposible dotarlo de un orden que restringido a $\mathbb{Z}$ coincida con el orden de $\mathbb{Z}$. El orden de $\mathbb{Z}$ no es lo que da la posibilidad de hablar del máximo común divisor de dos o más números, es la función valor absoluto. En el anillo $\mathbb{Z}[\omega]$, la función norma nos permitirá desarrollar una teoría del máximo común divisor en $\mathbb{Z}[\omega]$ tal como se realiza en $\mathbb{Z}$.

Un divisor en *común* de $z_1, z_2 \in \mathbb{Z}[\omega]$ es lo que esperamos; es un elemento $\gamma \in \mathbb{Z}[\omega]$, tal que $\gamma | z_1$ y $\gamma | z_2$. Por ejemplo, $(2-3\omega)|(5+2\omega)$ y $(2-3\omega)|(19+38\omega)$.

Teorema 4.1.5. *(Algoritmo de Euclides). Sean $z_1, z_2 \in \mathbb{Z}[\omega]$, con $z_2 \neq 0$. Entonces existe $\gamma \in \mathbb{Z}[\omega]$ tal que,*

1. *$\gamma | z_1$ y $\gamma | z_2$.*
2. *$\gamma = z_1\alpha_1 + z_2\alpha_2$, para algunos $\alpha_1, \alpha_2 \in \mathbb{Z}[\omega]$.*
3. *Si $\gamma' | z_1$ y $\gamma' | z_2$, entonces $\gamma' | \gamma$.*

Demostración. Aplicando repetidas veces el algoritmo de la división obtenemos,

$$\begin{aligned}
z_1 &= z_2k_1 + \delta_1 \quad y \quad \nu(\delta_1) < \nu(z_2), \\
z_2 &= \delta_1k_2 + \delta_1 \quad y \quad \nu(\delta_2) < \nu(z_1), \\
\delta_1 &= \delta_2k_3 + \delta_3 \quad y \quad \nu(\delta_3) < \nu(\delta_2), \\
&\vdots \qquad\qquad \vdots \\
\delta_{n-2} &= \delta_{n-1}k_n + \delta_n \quad y \quad \nu(\delta_n) < \nu(\delta_{n-1}), \\
\delta_{n-1} &= \delta_n k_{n+1} + 0.
\end{aligned}$$

Notemos que $\nu(z_2) > \nu(\delta_1) > \nu(\delta_2) > \cdots > \nu(\delta_n) > 0$ es una sucesión decreciente de enteros positivos y por lo tanto, en algún momento obtenemos un residuo $\delta_{n+1} = 0$. Queda claro que $\delta_n | \delta_{n-1}$, y por tanto, $\delta_n | \delta_{n-2}$. Continuando este proceso llegamos a que δ_n es un divisor común de z_1 y z_2. Por lo tanto, δ_n satisface la afirmación 1. Luego, despejando los δ_i para $i = 1, 2, \ldots, n$, en cada i-ésima igualdad y sustituyendo la anterior, se llega a

que δ_n cumple la afirmación 2.

Por último, si $\gamma^{'}|z_1, \gamma^{'}|z_2$, entonces $z_1 = \gamma^{'}t_1$ y $z_2 = \gamma^{'}t_2$, para algunos $t_1, t_2 \in \mathbb{Z}[\omega]$. Por la afirmación 2, tenemos que $\gamma = z_1\alpha_1 + z_2\alpha_2$, luego,

$$\gamma = z_1\alpha_1 + z_2\alpha_2 = \gamma^{'}t_1\alpha_1 + \gamma^{'}t_2\alpha_2 = \gamma^{'}(t_1\alpha_1 + t_2\alpha_2).$$

Por lo tanto, $\gamma^{'}|\gamma$. □

Un eiseniano γ que satisfaga el teorema anterior lo llamaremos *máximo común divisor* de z_1 y z_2 y lo denotaremos como (z_1, z_2) o bien $mcd(z_1, z_2)$.

La manera apropiada para definir dos enteros eisenianos primos relativos es la siguiente: supongamos que z_1 y z_2 son números eisenianos. Entonces, z_1 y z_2 son *primos relativos* si y solo si $(z_1, z_2) = u$, donde u es alguna unidad de $\mathbb{Z}[\omega]$.

Corolario 4.1.2. *Si $\gamma = (z_1, z_2)$ y $u \in U(\mathbb{Z}[\omega])$, entonces $u\gamma$ satisface el teorema 4.1.5.*

Demostración. Podemos suponer que γ satisface la propiedad 1 del teorema 4.1.5, es decir,

$$z_1 = \gamma t_1$$
$$z_2 = \gamma t_2.$$

Sea u una unidad de $\mathbb{Z}[\omega]$. Entonces $uz_1 = u\gamma t_1$ y $uz_2 = u\gamma t_2$, y como $U(\mathbb{Z}[\omega])$ es un grupo multiplicativo, entonces existe $u_2 \in U(\mathbb{Z}[\omega])$ tal que $u_2u = 1$. Así que $(u_2u)z_1 = u\gamma(u_2t_1)$ y $(u_2u)z_2 = u\gamma(u_2t_2)$, es decir, $u\gamma|z_1$ y $u\gamma|z_2$.

Para la afirmación 2 tenemos que $\gamma = z_1\alpha_1 + z_2\alpha_2$ y para algún elemento $u \in U(\mathbb{Z}[\omega])$, se sigue que,

$$u\gamma = uz_1\alpha_1 + uz_2\alpha_2 = z_1(u\alpha_1) + z_2(u\alpha_2).$$

Para la afirmación 3 tomamos en cuenta la afirmación 2 y suponemos que,

$$z_1 = \gamma^{'}t_1,$$
$$z_2 = \gamma^{'}t_2.$$

Luego,

$$\begin{aligned} u\gamma &= z_1(u\alpha_1) + z_2(u\alpha_2) \\ &= \gamma^{'}t_1(u\alpha_1) + \gamma^{'}t_2(u\alpha_2) \\ &= \gamma^{'}(t_1u\alpha_1 + t_2u\alpha_2). \end{aligned}$$

Por lo tanto, $\gamma^{'}|\gamma$. □

Una observación relevante del corolario anterior es que los enteros de Eisenstein z_1, z_2 tienen varios elementos que son su máximo común divisor y son exactamente seis. En este sentido cuando hacemos referencia a (z_1, z_2), nos referimos formalmente a cualquiera de los seis números que satisfacen el teorema 4.1.5. En la práctica queremos visualizar al (z_1, z_2) como un elemento único, así, por cuestiones prácticas tomaremos al elemento $1 \in U(\mathbb{Z}[\omega])$ como el máximo común divisor de z_1 y z_2, cuando z_1 y z_2 son primos relativos.

Ejemplo 4.1.4. *Calcule* $(4+\omega, 3-\omega)$.

Al aplicar varias veces el algoritmo de Euclides, observamos que,

$$\begin{aligned}
4+\omega &= (-\omega^2)(3-\omega)+(0-2\omega) && \text{y} \quad \nu(-2\omega) < \nu(3-\omega),\\
3-\omega &= (-\omega^2)(0-2\omega)+(1-\omega) && \text{y} \quad \nu(1-\omega) < \nu(-2\omega),\\
0-2\omega &= (1)(1-\omega)+(-1-\omega) && \text{y} \quad \nu(-1-\omega) < \nu(1-\omega),\\
1-\omega &= (\omega)(-1-\omega)+(-\omega) && \text{y} \quad \nu(-\omega) \leq \nu(-1-\omega),\\
-1-\omega &= (1)(-\omega)+(-1) && \text{y} \quad \nu(-1) \leq \nu(-\omega),\\
-\omega &= (\omega)(-1)+0,
\end{aligned}$$

por lo que, $(4+\omega, 3-\omega) = 1$*, por la convención del párrafo previo a este ejemplo.*

Teorema 4.1.6. *Sean* $z_1, z_2, \ldots, z_n \in \mathbb{Z}[\omega]$ *no todos cero. Existe* $\gamma \in \mathbb{Z}[\omega]$*, con las siguientes propiedades:*

1. $\gamma | z_i$*, para* $i = 1, 2, \ldots, n$.
2. *Si* $\gamma' | z_i$*, entonces* $\gamma' | \gamma$.

Demostración. Consideremos los conjuntos,

$$\begin{aligned}
A &= \{a_1 z_1 + a_2 z_2 + \cdots + a_n z_n : a_i \in \mathbb{Z}[\omega]\},\\
B &= \{\nu(x) : x \in A \setminus \{0\}\}.
\end{aligned}$$

Puesto que $B \cap \mathbb{N} \neq \emptyset$, entonces por el principio del buen orden existe $\gamma \in A$, de norma positiva mínima. Como $\gamma \in A$, entonces,

$$\gamma = a_1 z_1 + a_2 z_2 + \cdots + a_n z_n,$$

para algunos $a_1, a_2, \ldots, a_n \in \mathbb{Z}[\omega]$. Probaremos que para toda $x \in A$, $\gamma | x$. Para $x \in A$, tenemos que,

$$x = x_1 z_1 + x_2 z_2 + \cdots + x_n z_n.$$

Aplicando el algoritmo de la división tenemos,

$$x = \gamma k + r, \text{ con } \ 0 \leq \nu(r) < \nu(\gamma),$$

pero,

$$r = x - \gamma k = (x_1 - a_1 k) z_1 + (x_2 - a_2 k) z_2 + \cdots + (x_n - a_n k) z_n.$$

Así que, $r \in A$. Vemos que si $0 < \nu(r) < \nu(\gamma)$, entonces γ no es un elemento de A de norma positiva mínima. Por lo tanto, $\nu(r) = 0$ y $\gamma | x$. En particular si $z_i \in A$, se tiene que $\gamma | z_i$. Es claro que si $\gamma' | z_i$, entonces $\gamma' | \sum_{i=1}^{n} a_i z_i$ y por lo tanto, $\gamma' | \gamma$. □

El eiseniano γ del teorema anterior lo podemos llamar *máximo común divisor* de los eisenianos $z_1, z_2, \ldots, z_n$ y lo denotaremos como $mcd(z_1, z_2, \ldots, z_n)$ o bien $(z_1, z_2, \ldots, z_n)$.

Teorema 4.1.7. *Sean* $a, b, c \in \mathbb{Z}[\omega]$*. Se cumplen las siguientes afirmaciones:*

1. *Si* $a \neq 0$ *o* $b \neq 0$*, entonces la ecuación* $ax + by = c$ *tiene solución en los enteros eisenianos* x, y *si y solo si* $(a, b) | c$.
2. *(Teorema de Euclides). Si* $a | bc$ *y* $(a, b) = 1$*, entonces* $a | c$.

Demostración. Primero suponemos que la ecuación $ax + by = c$ tiene solución x, y en $\mathbb{Z}[\omega]$. Por la parte (1) del teorema 4.1.5 tenemos que $a = ga_1$, $b = ga_2$, donde $g = (a, b)$ y $a_1, a_2 \in \mathbb{Z}[\omega]$, luego,

$$ax + by = ga_1x + ga_2y = g(a_1x + a_2y) = c,$$

lo que implica que $g|c$.

Ahora, suponemos que $g|c$ y considerando la parte (2) del teorema 4.1.5, podemos escribir $g = ak_0 + bl_0$, para algunos k_0, l_0 en $\mathbb{Z}[\omega]$. Por lo tanto,

$$c = gt = (ak_0 + bl_0)t = a(k_0t) + b(l_0t), \quad \text{para algún } t \in \mathbb{Z}[\omega].$$

Para la segunda afirmación del teorema, tenemos que, por hipótesis $bc = at_1$, para algún $t_1 \in \mathbb{Z}[\omega]$ y por la parte (1) del teorema 4.1.5 tenemos que $1 = ax+by$, para algunos $x, y \in \mathbb{Z}[\omega]$, luego,

$$\begin{aligned} 1 &= ax + by, \\ c &= cax + cby, \qquad \textit{al multiplicar por } c, \\ c &= acx + bcy, \\ c &= cax + at_1y, \qquad \textit{al sustituir el factor } bc, \\ c &= a(cx + t_1y). \end{aligned}$$

Por lo tanto, $a|c$. □

Supongamos que $z_1, z_2 \in \mathbb{Z}[\omega]\setminus\{0\}$. Un *múltiplo común* de z_1, z_2 es un entero eiseniano γ, tal que $z_1|\gamma$ y $z_2|\gamma$. Es claro que z_1z_2 es múltiplo común de z_1 y z_2. En general si $z_1, z_2, \ldots, z_n \in \mathbb{Z}[\omega] \setminus \{0\}$, el conjunto,

$$M = \{x \in \mathbb{Z}[\omega] \setminus \{0\} : z_i|x, i = 1, 2, \ldots, n\},$$

es distinto del vacío $\emptyset$ pues $\prod_{i=1}^{n} z_i \in M$. Con lo anterior hemos justificado que al menos existe un múltiplo común de $z_1, z_2, \ldots, z_n$.

Teorema 4.1.8. *Sean $z_1, z_2, \ldots, z_n \in \mathbb{Z}[\omega] \setminus \{0\}$, y defina el conjunto M como antes. Existe $m \in \mathbb{Z}[\omega]$ con las siguientes propiedades:*

1. *$z_i|m$, para $i = 1, 2, \ldots, n$.*
2. *Si $m' \in M$, entonces $m|m'$.*

Demostración. Sea $H = \{\nu(x) : x \in M\}$. Notemos que $\emptyset \neq H \subseteq \mathbb{N}$. Por el principio del buen orden existe $h \in H$ de norma mínima. Así, $h = \nu(m)$, para algún $m \in M$. El elemento m satisface la afirmación 1, ya que pertenece al conjunto H.

Sea $m' \in M$. Al aplicar el algoritmo de la división tenemos que,

$$m' = km + r, \text{ con } \ 0 \leq \nu(r) < \nu(m).$$

Si $\nu(r) \neq 0$, entonces $r = m' - km \in M$, lo cual es absurdo por la elección de m. Por lo tanto, $r = 0$ y $m|m'$. □

El eiseniano m del teorema anterior lo podemos llamar *mínimo común múltiplo* de $z_1, z_2, \ldots, z_n$ y lo denotamos como $mcm(z_1, z_2, \ldots, z_n)$.

4.2. Factorización única en $\mathbb{Z}[\omega]$

Cualquier par de enteros eisenianos z_1, z_2 tiene por lo menos seis divisores en común,

$$1, -1, \omega, -\omega, \omega^2, -\omega^2.$$

Si z_1, z_2 solo comparten estos seis divisores, entonces es claro que z_1 y z_2 son primos relativos.

Si $z \neq 0$ no es asociado de 1, entonces z también admite como divisores a sus asociados. Por lo tanto, cada entero eiseniano z tiene al menos,

$$1, -1, \omega, -\omega, z, -z, \omega z, -\omega z, \omega^2 z, -\omega^2 z,$$

como divisores. En analogía con el caso de los enteros $\mathbb{Z}$, recordemos que un número primo $p \in \mathbb{Z}$ es aquel que tiene exactamente cuatro divisores,

$$1, -1, p, -p.$$

4.2.1. Números primos en $\mathbb{Z}[\omega]$

En algunos de los problemas importantes en la teoría de números, destaca la factorización de un entero racional. Si se conociera una lista completa de números primos racionales, seguramente podríamos factorizar cualquier entero racional. Afortunadamente no es así, de hecho no se sabe cómo generarlos a todos. Lo que sí se sabe es que no existe, de forma eficiente, un polinomio en una variable que reproduzca sólo primos racionales. Al respecto existen muchas conjeturas, por ejemplo: existe una infinidad de números primos racionales de la forma $x^2 + 1$. Para empezar a resolver el problema de factorizar un entero de Eisenstein, primero analizaremos de manera explícita a los primos de Eisenstein.

En analogía con $\mathbb{Z}$, si $\pi \in \mathbb{Z}[\omega] \setminus \{0\}$ no es unidad y solo admite como divisores a sus unidades y a sus asociados, entonces diremos que π es *primo*. Para evitar confusiones en el lenguaje, llamaremos *primos racionales* a los números primos de $\mathbb{Z}$ y simplemente *primos*, a los de $\mathbb{Z}[\omega]$. El siguiente resultado es consecuencia de la definición de número primo.

Lema 4.2.1. *Sea π un número eiseniano. Si $\nu(\pi) > 1$ y π no es producto de enteros eisenianos de norma mayor que 1, entonces π es primo.*

Demostración. La hipótesis π no es producto de enteros eisenianos de norma mayor que 1 significa que en cualquier factorización de π, al menos uno de sus factores es una unidad. Así que $\pi = u\beta$, para algún $u \in U(\mathbb{Z}[\omega])$. Por tanto, $\beta|\pi$ y $\beta = u^{-1}\pi$, lo que además significa que β es un asociado de π. Se concluye que π es primo. □

Una interpretación del lema anterior es que si $\nu(\pi) > 1$ y $\pi = \alpha\beta$ implica que $\nu(\alpha) = 1$ o $\nu(\beta) = 1$, entonces π es primo.

Uno de los objetivos de esta sección es identificar a los primos en $\mathbb{Z}[\omega]$ y dar un método para factorizar enteros eisenianos. El siguiente resultado proporciona un método elemental para identificar algunos de ellos.

Teorema 4.2.1. *Sea $z \in \mathbb{Z}[\omega]$, tal que $\nu(z)$ es un primo racional. Entonces z es primo.*

Demostración. Supongamos que $\nu(z) = p$, con p un primo racional y $z = z_1 z_2$. Entonces,

$$p = \nu(z) = \nu(z_1)\nu(z_2).$$

Así que $\nu(z_1) = 1$ o $\nu(z_2) = 1$ y por tanto, z_1 o z_2 es unidad. Se sigue que z es primo. □

Ejemplo 4.2.1. *El número eiseniano* $3 + 2\omega$ *es primo, pues* $\nu(3 + 2\omega) = 7$.

Ejemplo 4.2.2. *El número eiseniano* $1 + 2\omega$ *es primo, pues* $\nu(1 + 2\omega) = 3$.

Ejemplo 4.2.3. *El número eiseniano* $5 + 3\omega$ *es primo, pues* $\nu(5 + 3\omega) = 19$.

Lema 4.2.2. *Cada entero eiseniano* z*, con* $\nu(z) > 1$ *tiene una representación como producto finito de primos.*

Demostración. Por inducción sobre $\nu(z)$. Si $\nu(z) = 2$, entonces por el teorema 4.2.1 se tiene que z es primo y así el lema queda demostrado. Supongamos el lema válido para cualquier eiseniano α, con $2 \leq \nu(\alpha) < \nu(z)$. Si z es primo, entonces concluimos la demostración del lema, pues z es en sí un primo. Si z no es primo, entonces $z = z_1 z_2$, con

$$1 < \nu(z_1) < \nu(z) \quad \text{y} \quad 1 < \nu(z_2) < \nu(z).$$

Ahora, por hipótesis de inducción tenemos que,

$$z_1 = \pi_1 \pi_2 \cdots \pi_r, \quad \text{y} \quad z_2 = \pi_1' \pi_2' \cdots \pi_t',$$

donde los π_i, π_j' son primos. Por lo tanto, $z = \pi_1 \pi_2 \cdots \pi_r \cdot \pi_1' \pi_2' \cdots \pi_t'$. □

Por definición, cualquier primo π tiene como únicos divisores a,

$$1, -1, \omega, -\omega, \pi, -\pi, \omega\pi, -\omega\pi, \omega^2\pi, -\omega^2\pi.$$

Así, $p \in \mathbb{Z}[\omega]$ es primo si tiene exactamente diez divisores.

El siguiente resultado muestra que los primos racionales son primos eisenianos.

Lema 4.2.3. *Sea* $n \in \mathbb{Z}[\omega]$ *un primo. Entonces* n *es un primo racional.*

Demostración. Si $n \in \mathbb{Z}[\omega]$ es primo, entonces sus únicos divisores son:

$$1, -1, \omega, -\omega, n, -n, \omega n, -\omega n, \omega^2 n, -\omega^2 n.$$

El resultado se sigue de observar que los únicos divisores racionales (es decir números enteros racionales) de n son,

$$1, -1, n, -n.$$

□

Lema 4.2.4. *Si* π *es primo y* $\pi | ab$*, entonces* $\pi | a$ *o* $\pi | b$.

Demostración. Supongamos que $\pi \nmid a$. Si $\gamma = (\pi, a)$, entonces γ es alguno de los elementos del conjunto $\{\pm 1, \pm\omega, \pm\pi, \pm\omega\pi, \pm\omega^2\pi\}$ y así $\gamma \in U(\mathbb{Z}[\omega])$. Por la afirmación 2 del teorema 4.1.7 concluimos que $\pi | b$. □

Como caso particular del lema anterior, si π_1, π_2, π_3 son primos y $\pi_1 | \pi_2 \pi_3$, entonces π_1 es asociado de π_2 o π_3. Esto es fundamental para el siguiente resultado.

Teorema 4.2.2. *(Teorema Fundamental de la Aritmética en $\mathbb{Z}[\omega]$). La representación de un entero eiseniano $z \neq 0$, y distinto de una unidad, como producto finito de primos es única salvo el orden de los primos y asociados.*

Demostración. Sea $z \in \mathbb{Z}[\omega]$, con $\nu(z) > 1$. Supongamos que z tiene dos factorizaciones:

$$z = \pi_1\pi_2 \cdots \pi_r = \pi'_1\pi'_2 \cdots \pi'_t,$$

con $r < t$. De la ecuación anterior se observa que $\pi_1 | \pi'_1\pi'_2 \cdots \pi'_t$ y por tanto π_1 es un asociado de algún π'_j. Sin pérdida de generalidad podemos suponer que π_1 es asociado de π'_1. Cancelando π_1 y π'_1 obtenemos que,

$$u_1\pi_2 \cdots \pi_r = \pi'_2\pi'_3 \cdots \pi'_t,$$

donde u_1 es alguna unidad. Repetimos el argumento anterior r-veces para llegar finalmente a que,

$$u_1u_2 \cdots u_r = \pi'_{r+1}\pi'_{r+2} \cdots \pi'_t.$$

Por consiguiente,

$$1 = \nu(u_1)\nu(u_2) \cdots \nu(u_r) = \nu(\pi'_{r+1})\nu(\pi'_{r+2}) \cdots \nu(\pi'_t),$$

lo cual es absurdo, pues $\nu(\pi'_j) > 1$. Por lo tanto, $r \geq t$. Si ahora suponemos que $t < r$, usamos el mismo argumento anterior y obtenemos un absurdo. Se sigue que $t = r$, y en consecuencia se tiene la factorización única de z. □

Ejemplo 4.2.4. *Sea $z = 2 + 4\omega$. La factorización de z como producto de primos es la siguiente:*

$$2 + 4\omega = (2 + 0\omega)(1 + 2\omega).$$

Ejemplo 4.2.5. *Sea $z = 4 - \omega$. Entonces z se factoriza como producto de los siguientes enteros eisenianos primos:*

$$4 - \omega = (2 + \omega)(1 - 2\omega).$$

Proposición 4.2.1. *Si π es un primo, existe un primo racional p, tal que $\nu(\pi) = p$ o $\nu(\pi) = p^2$. Además, en el caso que $\nu(\pi) = p$, entonces π no es asociado a un primo racional y en caso que $\nu(\pi) = p^2$, entonces π es asociado del primo racional p.*

Demostración. Sea π un primo. Por definición de norma $\nu(\pi) = \pi\bar{\pi} = n$, para algún entero racional positivo n, y como π no es unidad, se tiene que $n > 1$.

Ahora, como $n > 1$, entonces podemos escribir a n como producto de primos racionales, digamos $n = \pi\bar{\pi} = p_1^{n_1}p_2^{n_2} \ldots p_k^{n_k}$, lo cual implica que algún p_i es tal que $\pi | p_i$, para $1 \leq i \leq k$.

Supongamos que $1 \leq i \leq k$ es fijo y que $p_i = \pi\gamma$, para algún $\gamma \in \mathbb{Z}[\omega]$. Entonces $\nu(\pi)\nu(\gamma) = \nu(\pi\gamma) = \nu(p_i) = p_i^2$. Por lo tanto, $\nu(\pi) = p_i^2$ y $\nu(\gamma) = 1$ o bien $\nu(\pi) = p_i$ y $\nu(\gamma) = p_i$. En el caso que $\nu(\pi) = p_i^2$ y $\nu(\gamma) = 1$, tenemos que γ es unidad y por lo tanto, tiene inverso γ^{-1} en $\mathbb{Z}[\omega]$, el cual también es unidad. Así, $\pi = \gamma^{-1}p_i$. Se sigue que π es asociado de algún primo racional.

Ahora, consideramos el caso en que $\nu(\pi) = p_i$ y $\nu(\gamma) = p_i$. Si suponemos que π es asociado de algún primo racional q, tenemos que $\pi = \mu q$, con $\mu \in \mathbb{Z}[\omega]$. Entonces vemos

que $p_i = \nu(\pi) = \nu(\mu)\nu(q) = q^2$, lo cual es una contradicción con el hecho de que p_i es primo racional. Por lo tanto, cuando $\nu(\pi) = p_i$, concluimos que π no es asociado de algún primo racional. □

Proposición 4.2.2. *Supongamos que π es primo. Entonces $\nu(\pi) = 3$ si y solo si se cumple que $\pi = \mu(1-\omega)$, para alguna unidad μ.*

Demostración. Escribimos $\gamma = 1-\omega$. Si $\pi = \mu\gamma$, para alguna unidad μ, entonces tenemos que $\nu(\pi) = \nu(\mu)\nu(\gamma) = 3$.
Supongamos que $\nu(\pi) = 3$ y vamos probar que $\pi = \mu\gamma$. Sea $\pi = a + b\omega$. Como $\nu(\pi) = 3$, podemos escribir,

$$a^2 - ab + b^2 = 3, \tag{4.1}$$

es decir, $a^2+b^2-2ab+ab = 3$, simplificando esta ecuación obtenemos que $(a-b)^2 = 3-ab$. Como $a^2 + b^2 \geq 0$, tenemos que $3 + ab \geq 0$. Es decir, $-ab \leq 3$, y dado que $(a-b)^2 \geq 0$, se sigue que $3 - ab \geq 0$, de modo que $ab \leq 3$; por lo tanto, obtenemos que $-3 \leq ab \leq 3$.

A continuación analizamos los siguientes tres casos: los números a y b tienen signos opuestos o a y b tienen el mismo signo o a y b son cero.

Consideramos el caso en que a y b tienen signos opuestos. Supongamos que $a < 0$ y $b > 0$. Las únicas soluciones enteras para $ab \geq -3$ son $a = -3$ y $b = 1$ o bien $a = -1$ y $b \in \{1, 2, 3\}$. La solución $a = -1$ y $b = 1$ satisface la ecuación (4.1). Por lo tanto, $\pi = -1 + \omega = (-1)\gamma$. Por simetría, la única solución para $ab \geq -3$ con $a > 0$ y $b < 0$ que satisface la ecuación (4.1) es $a = 1$ y $b = -1$, así que $\pi = 1 - \omega = (1)\omega$.

Ahora, consideramos el caso en que a y b tienen el mismo signo. Aquí, $a > 0$ y $b > 0$. Las únicas soluciones enteras que satisfacen $ab \leq 3$ son $a = 1$ y $b \in \{1, 2, 3\}$ o bien $b = 1$ y $a \in \{1, 2, 3\}$. De estas soluciones solo $\pi = 2 + \omega = -\omega^2\gamma$ y $\pi = 1 + 2\omega = \omega\gamma$ satisfacen la ecuación (4.1).
Por simetría, dado que $\nu(\pi) = \nu(-\pi)$, tenemos que,

$$\pi = -2 - \omega = \omega^2\pi \quad \text{y} \quad \pi = -1 - 2\omega = -\omega\pi,$$

son las únicas soluciones para $a < 0$ y $b < 0$.

Finalmente consideramos el caso cuando $a = 0$ o $b = 0$. Si $a = 0$, la ecuación (4.1) se escribe como $b^2 = 3$, la cual no tiene solución entera. Similarmente si $b = 0$, la ecuación (4.1) se escribe como $a^2 = 3$, la cual no tiene solución entera.

Por lo tanto, las únicas soluciones para $\nu(\pi) = 3$ son $\pi = (\pm 1)\gamma$, $\pi = \pm\omega\gamma$ y $\pi = \pm\omega^2\gamma$. Se concluye que $\nu(\pi) = 3$ si y solo si $\pi = \mu(1-\omega)$, para alguna unidad μ. □

Los siguientes teoremas nos proporcionan una clasificación de los primos de Eisenstein.

Teorema 4.2.3. *Supongamos que p es un primo racional. Entonces las siguientes afirmaciones son ciertas:*

1. *$1 - \omega$ y sus asociados son primos de Eisenstein.*
2. *Si $p \equiv 2 \pmod 3$, entonces p y sus asociados son primos de Eisenstein.*

3. Si $p \equiv 1 \pmod 3$, entonces $p = \pi\bar{\pi}$, donde π es un primo.

Demostración. Vemos que $\nu(1-\omega) = 3$, y por el teorema 4.2.1 se sigue que $1-\omega$ es primo. Es claro que sus asociados también son primos, pues son múltiplos de $1-\omega$ con norma igual a 3.
Sea $p \neq 3$ un primo racional. Verificaremos la siguiente condición: p es primo si y solo si $p \equiv 2 \pmod 3$.

Si p no es primo, entonces $p = \alpha\beta$, con $\alpha, \beta \in \mathbb{Z}[\omega]$ y $\nu(\alpha) > 1$, $\nu(\beta) > 1$. Luego, vemos que $p^2 = \nu(p) = \nu(\alpha)\nu(\beta)$, así que $\nu(\alpha) = p$ y $\nu(\beta) = p$.

Sea $\alpha = a + b\omega \in \mathbb{Z}[\omega]$. Entonces $p = \nu(\alpha) = a^2 + b^2 - ab$, es decir, $(a+b)^2 - 3ab = p$. Por lo tanto, $p \equiv (a+b)^2 \pmod 3$, lo cual implica que $p \equiv 1 \pmod 3$, ya que 1 es el único cuadrado distinto de cero módulo 3.

Recordemos que x es un residuo cuadrático si existe un entero z, tal que se cumple la congruencia $z^2 \equiv x \pmod y$. De ley de reciprocidad cuadrática recordamos un corolario, el cual establece lo siguiente: supongamos que p y q son primos racionales impares y distintos, con $q \equiv 1 \pmod 4$. Entonces q es un cuadrado en el campo $\mathbb{F}_p$ si y solo si p es un cuadrado en $\mathbb{F}_q$. [[32], capítulo 3, corolario 3.2.9].

Supongamos que en la relación $1 \equiv p \pmod 3$ el número p es un residuo cuadrático módulo 3. Entonces $p = 1 + y(3)$, para algún $y \in \mathbb{Z}$, es decir, $p = 1 - y(-3)$ y $p \equiv 1 \pmod{-3}$, así que p es un residuo cuadrático módulo -3. En consecuencia, por el corolario descrito anteriormente tomamos $q = -3$ y obtenemos que $q = -3 \equiv 1 \pmod 4$; se sigue que -3 es un residuo cuadrático módulo p. Esto quiere decir que existe $a \in \mathbb{Z}$ tal que $a^2 \equiv -3 \pmod p$ o bien $a^2 = -3 + bp$, para algún $b \in \mathbb{Z}$.

Considerando que $\sqrt{-3} = 1 + 2\omega$, tenemos que:

$$a^2 + 3 = (a - \sqrt{-3})(a + \sqrt{-3}) = (a - 1 - 2\omega)(a + 1 + 2\omega).$$

Ahora, como p divide a $a^2 + 3$ y si suponemos que p es primo, entonces se tiene que p divide a alguno de los factores de $a^2 + 3$. Además, para cada $p \in \mathbb{Z}$ y $p|c + d\omega$, se cumple que $p|c$ y $p|d$.

Si p fuera primo, entonces tendría que dividir uno de los factores de $a^2 + 3$, pero esto no puede suceder ya que $p \equiv 1 \pmod{-3}$, es decir, $p \neq 2$. Por lo tanto, $p \nmid 2$ y además p no divide a alguno de los factores de $a^2 + 3$, lo que significa que p no es primo.

Por lo tanto, un primo racional $p \neq 3$ no es primo si y solo si $p \equiv 1 \pmod 3$. Luego, concluimos que $p \neq 3$ es primo si y solo si $p \equiv 2 \pmod 3$, y así queda probado el enunciado número 2 de este teorema.

Si $p \equiv 1 \pmod 3$, entonces p no es primo y $p = \pi\beta$, donde $\pi, \beta \in \mathbb{Z}[\omega]$ no son unidades, luego, vemos que $\nu(p) = \nu(\pi)\nu(\beta) = p^2$. Se sigue que $\nu(\pi) = p$, y por el teorema 4.2.1 se concluye que π es primo. □

Teorema 4.2.4. *Si π es primo, entonces se satisface una de las siguientes tres condiciones:*

1. *El número eiseniano π es asociado de $1-\omega$.*
2. *El número eiseniano π es asociado de un primo racional p que satisface la congruencia $p \equiv 2$ (mód 3).*
3. *El valor $\nu(\pi)$ es un primo racional y satisface $p \equiv 1$ (mód 3).*

Demostración. Sea π un primo. Por la Proposición 4.2.1 podemos suponer que $\nu(\pi) = p$, para algún primo racional p. Por el Teorema 4.2.3 sucede alguno de los siguientes casos: $p \equiv 1$ (mód 3) o $p \equiv 2$ (mód 3) o $p = 3$.

Si $p = 3$, entonces π satisface la Proposición 4.2.2, es decir, π es asociado de $(1-\omega)$, y por tanto π cae en la primer categoría del teorema.

Si $p \equiv 2$ (m*ód* 3), entonces por el Teorema 4.2.3 se sigue que p es primo, pero también vemos que $p = \nu(\pi) = \pi\bar{\pi}$, se tiene una contradicción. Por lo tanto, p satisface la condición $p \equiv 1$ (mód 3), cumpliéndose así la tercera condición del teorema.

Por la Proposición 4.2.1 podemos suponer que $\nu(\pi) = p^2$, con p un primo racional, además esta proposición nos asegura que π es un asociado de p. Por lo tanto, p es primo. Observemos que $p = 3$ no es primo, pues de la relación $\sqrt{-3} = 1 + 2\omega$ obtenemos $3 = -(1+2\omega)^2$ y $(1+2\omega)$ no es unidad. Luego, $p \neq 3$ es primo si y solo si $p \equiv 2$ (mód 3). Por lo tanto, $\nu(\pi) = p^2$, es decir, π cae en la categoría número dos del teorema. □

Proposición 4.2.3. *Sea π un primo, entonces hay seis o doce números primos con la misma norma que π.*

Demostración. Sea π un primo. Entonces π está en una de las tres categorías del teorema 4.2.4. Supongamos que π está en la primera categoría. Entonces $\nu(\pi) = 3$ y por la proposición 4.2.2 se sigue que π tiene seis asociados de norma 3.
Ahora, si π está en la segunda categoría, tenemos que π es asociado de un primo racional p, tal que $p \equiv 2$ (mód 3) y $\nu(\pi) = p^2$. Por la Proposición 4.2.1 se deduce que si $\alpha \in \mathbb{Z}[\omega]$ satisface que $\nu(\alpha) = p^2$, donde p es un primo racional, entonces α es un asociado de p y por lo tanto, asociado de π. Luego, los seis asociados de p son los únicos números primos con norma p^2.

Finalmente si π está en la tercera categoría, tenemos que $\nu(\pi) = p$, con p un primo racional, tal que $p \equiv 1$ (mód 3). Además, $\pi\bar{\pi} = \nu(\pi) = p$. Por el Teorema Fundamental de la Aritmética en $\mathbb{Z}[\omega]$, tenemos que la factorización anterior es única salvo asociados. Por lo tanto, los asociados de π y de $\bar{\pi}$ son los únicos primos que tienen norma p. En el caso donde $\bar{\pi}$ no es un asociado de π, se tendrán doce primos con norma p, y en caso contrario habrá seis primos con norma p. □

Lema 4.2.5. *Sea $\alpha = a + b\omega \in \mathbb{Z}[\omega]$. Si $d = (a, b)$, entonces $d^2 | \nu(\alpha)$.*

Demostración. Como $d = (a, b)$, entonces $a = dt_1, b = dt_2$, para algunos $t_1, t_2 \in \mathbb{Z}$. Por definición $\nu(\alpha) = a^2 + b^2 - ab$. Luego, vemos que,

$$\begin{aligned} a^2 + b^2 - ab &= (dt_1)^2 + (dt_2)^2 - ab \\ &= d^2t_1^2 + d^2t_2^2 - (dt_1)(dt_2) \\ &= d^2(t_1^2 + t_2^2 - t_1t_2) \\ &= d^2t, \end{aligned}$$

donde $t = t_1^2 + t_2^2 - t_1 t_2$ es un número entero. Por lo tanto, $d^2 | \nu(\alpha)$. □

El siguiente resultado expresa cómo construir un entero pequeño en el lattice generado por un entero eiseniano.

Lema 4.2.6. *Sea $\alpha = a + b\omega \in \mathbb{Z}[\omega]$ distinto de cero. Si $d = (a, b)$, entonces $\dfrac{\nu(\alpha)}{d} \in L(\alpha)$ y $\dfrac{\nu(\alpha)}{d}$ es el entero positivo más pequeño en $L(\alpha)$.*

Demostración. Recordemos que $L(\alpha) = \{\gamma\alpha | \gamma \in \mathbb{Z}[\omega]\}$. Por hipótesis $d = (a, b)$, $b = dt_1$ y $a = dt_2$, con $t_1, t_2 \in \mathbb{Z}$. También notemos que,

$$a - b = dt_2 - dt_1 = d(t_2 - t_1) \quad \text{y} \quad t_2 - t_1 \in \mathbb{Z}.$$

Así, concluimos que $d | a - b$. Luego,

$$\frac{\nu(\alpha)}{d} = \frac{\alpha\bar{\alpha}}{d} = \alpha\frac{\bar{\alpha}}{d} = \alpha\left[\frac{(a-b) - b\omega}{d}\right] = \alpha\left[\frac{dt_3}{d} + \frac{dt_1}{d}\omega\right] = \alpha[t_3 - t_1\omega] \in L(\alpha).$$

Supongamos que $\lambda\bar{\alpha} = \lambda(a - b - b\omega) \in \mathbb{Z}[\omega]$, con $\lambda \in \mathbb{R}$ y $\lambda \neq 0$. Se sigue que $\lambda(a - b), \lambda b \in \mathbb{Z}$. Por lo tanto, λ es un número racional de la forma $\dfrac{m}{n}$, donde m y n son enteros racionales, primos relativos tal que $n | a - b$, $n | b$.

Sea $y \in L(\alpha) \cap \mathbb{Z}$. Entonces $y = \lambda\bar{\alpha}\alpha$, con $\lambda\bar{\alpha} \in \mathbb{Z}[\omega]$, para algún $\lambda \in \mathbb{R}$. Sabemos que $d = (a, b) = (a - b, b)$ y d es el entero más grande que divide a $\bar{\alpha}$. Por lo tanto, $\lambda = 1/d$ es el valor positivo más pequeño para λ.
Por lo tanto, $y = \dfrac{\alpha\bar{\alpha}}{d} = \dfrac{\nu(\alpha)}{d}$ es el entero positivo más pequeño en $L(\alpha)$. □

Enseguida se calcula la cardinalidad del cociente $\mathbb{Z}[\omega]/(\alpha)$.

Teorema 4.2.5. *Sea $\alpha \in \mathbb{Z}[\omega]$ distinto de cero. Se tienen exactamente $\nu(\alpha)$ elementos en $\mathbb{Z}[\omega]/(\alpha)$. Esto es, se tienen exactamente $\nu(\alpha)$ clases de congruencia módulo α.*

Demostración. Sea $\alpha = a + b\omega \in \mathbb{Z}[\omega]$ y sea $d = (a, b)$. Por el lema 4.2.6, $x = \dfrac{\nu(\alpha)}{d} \in \mathbb{Z}$. Verificaremos que,

$$C = \{u + v\omega | u, v \in \mathbb{Z}, 0 \leq u < x, 0 \leq v < d\},$$

es un conjunto completo de representantes de clases de congruencia módulo α y con esto probaremos que hay exactamente $xd = \nu(\alpha)$ clases de congruencia.

Sea $\beta = m + n\omega \in \mathbb{Z}[\omega]$. Podemos escribir $n \equiv v \pmod{d}$, con $0 \leq v < d$, es decir, $n = v + kd$, para algún $k \in \mathbb{Z}$. Como $d = (a, b)$, entonces existen enteros s, t tales que $d = sb + ta$ y así tenemos las siguientes implicaciones,

$$ksb + kta = kd \Rightarrow -ksb - kta = -kd \Rightarrow n - ksb - kta = n - kd = v.$$

También, podemos escribir $\beta - \gamma \equiv \beta \pmod{\alpha}$, para algún $\gamma \in L(\alpha)$. El objetivo es encontrar $\gamma \in L(\alpha)$ tal que $\beta - \gamma = u + v\omega$, con $0 \leq u < x$ y $0 \leq v < d$. Vemos que:

$$\begin{aligned} \beta - ks\alpha + kt\omega^2\alpha &= (m + n\omega) - ks(a + b\omega) + kt(-a + b - a\omega) \\ &= (m - ksa - kta + ktb) + (n - ksb - kta)\omega \\ &= (m - ksa - kta + ktb) + v\omega. \end{aligned}$$

Por otro lado, tenemos que $m - ksa - kta + ktb \equiv u \pmod{x}$, con $0 \leq u < x$. Por lo tanto, se cumple la relación $m - ksa - kta + ktb = u + ax$, para algún $s \in \mathbb{Z}$, y como $\alpha, \omega^2, x \in L(\alpha)$, se sigue que,

$$\beta - ks\alpha + kt\omega^2\alpha - sx = (m - ksa - kta + ktb - sx) + (n - ksb - kta)\omega = u + v\omega,$$

con $0 \leq u < x$, $0 \leq v < d$.

Hemos visto que todo número de Eisenstein es congruente módulo α con un entero de Eisenstein de la forma $u + v\omega$, con $0 \leq u < x$ y $0 \leq v < d$. Falta probar que el conjunto C es un conjunto único de clases de congruencia. Sea $\beta = m + n\omega$ y $\beta' = m' + n'\omega$, con $0 \leq m, m' < x$ y $0 \leq n, n' < d$ y supongamos que $\beta \equiv \beta' \pmod{\alpha}$, $\beta - \beta' = \alpha\gamma$, para algún $\gamma \in \mathbb{Z}[\omega]$. Dado que $\alpha = a + b\omega$ y $d = (a, b)$, tenemos que $\dfrac{\alpha}{d} \in \mathbb{Z}[\omega]$. Por lo tanto,

$$\frac{(m - m') + (n - n')\omega}{d} = \frac{\beta - \beta'}{d} = \frac{\alpha\gamma}{d} = \frac{\alpha}{d}\gamma \in \mathbb{Z}[\omega],$$

de modo que $d|(m - m')$ y $d|(n - n')$. Como $0 \leq n - n' < d$ y $d|n - n'$, entonces $n - n' = 0$, es decir, $n = n'$. Por lo tanto, tenemos que $m - m' = \alpha\gamma$ y en consecuencia $\alpha\gamma \in \mathbb{Z}[\omega]$. Sabemos que x es el número entero positivo más pequeño en $L(\alpha)$, y como $0 \leq m - m' < x$, se sigue que $m - m' = 0$ o equivalentemente $m = m'$. Por lo tanto, C es un conjunto completo de representantes para las clases de congruencia módulo α y hay exactamente $xd = \nu(\alpha)$ distintas clases de congruencia. □

Capítulo 5

Criptosistema ETRU

En el capítulo 4 se estudió el anillo de los enteros de Eisenstein y sus propiedades, algunos de los conceptos vistos en este capítulo son analogías de las definiciones vistas en [29], [19] y [10], que fueron adaptadas por K. Jarvis al anillo $\mathbb{Z}[\omega]$, y que serán útiles para implementar el cifrado NTRU sobre dicho anillo, con una notación definida por J. Silverman en [10]. Veremos algunos de los beneficios que tiene este cifrado en comparación con el cifrado NTRU y se calculará la probabilidad de fallo de descifrado basado en el modelo que propone K. Jarvis en [20]. También se estudia un ataque común a este cifrado, el ataque mediante lattices, dando un desarrollo explícito del análisis de la probabilidad de fallo y el ataque mediante lattices con una notación estándar.

5.1. Criptosistema NTRU sobre los enteros de Eisenstein

Escribiremos ETRU para referirnos al cifrado NTRU sobre el anillo de los enteros de Eisenstein. Definimos los anillos base que ocupa este cifrado: $R = \mathbb{Z}[\omega][x]/\langle x^n - 1\rangle$ es el anillo de polinomios módulo el ideal generado por el polinomio $x^n - 1$, con coeficientes en el anillo de los enteros de Eisenstein, $R_p = \mathbb{Z}[\omega]_p[x]/\langle x^n - 1\rangle$ y $R_q = \mathbb{Z}[\omega]_q[x]/\langle x^n - 1\rangle$ son los anillos de polinomios módulo el ideal generado por el polinomio $x^n - 1$ con coeficientes en los anillos de enteros de Eisenstein módulo p y q respectivamente.

La manera de cifrar un mensaje con ETRU es similar a la del cifrado NTRU. A continuación definimos algunos conceptos útiles.

Definición 5.1.1. *Un ω-levantamiento de un polinomio $f(x) \in R_q$ a R es el único polinomio $f'(x) \in R$ que satisface $f'(x) \mod q = f(x)$.*

En el corolario 4.1.1 del capítulo 4 vimos que las unidades en $\mathbb{Z}[\omega]$ son $\pm 1, \pm\omega, \pm\omega^2$, y como sabemos, estos elementos son invertibles y se consideran en la siguiente definición.

Definición 5.1.2. *Definimos el conjunto de polinomios ternarios en $\mathbb{Z}[\omega]$ como sigue,*

$$T(d_1, d_2) = \left\{ a(x) \in R : \begin{array}{c} a(x) \text{ tiene } d_1 \text{ coeficientes iguales a } 1,\ d_2 \text{ coeficientes} \\ \text{de cada una del resto de las unidades y el resto} \\ \text{de sus coeficientes iguales a } 0. \end{array} \right\}$$

Una vez que tenemos presentes los conceptos anteriores se eligen parámetros (n, p, q, d), tales que n es un primo racional, p y $q \in \mathbb{Z}[\omega]$ son primos relativos con la condición

$(n, q) = (p, q) = 1$ y d es un entero racional positivo. Para fines prácticos se suele tomar a p y q como primos, pero esto no es indispensable en el cifrado.

5.1.1. Generación de claves

Para cifrar un mensaje con ETRU, Aldo elige parámetros (n, p, q, d) que satisfagan las condiciones mencionadas anteriormente. Luego, selecciona de manera aleatoria dos polinomios $f(x) \in T(d+1, d)$ y $g(x) \in T(d, d)$ tal que el polinomio $x-1$ no sea un factor de $f(x)$, para así reducir la posibilidad de que $f(x)$ no sea invertible.

Aldo encuentra los polinomios $F_p(x) = f^{-1}(x) \in R_p$ y $F_q(x) = f^{-1}(x) \in R_q$, es decir, el polinomio inverso de $f(x)$ en R_p y R_q respectivamente.

En caso que el polinomio $f(x)$ no tenga inverso en R_p o en R_q, Aldo debe seleccionar otro polinomio que sí tenga inverso multiplicativo en R_p y en R_q respectivamente. Enseguida se calcula $h(x) = F_q(x)g(x) \mod q$.

Entonces las claves de Aldo para ETRU son:

1. Se deja como clave pública al polinomio $h(x)$.
2. Los parámetros n, p, q también son públicos.
3. La clave privada de Aldo serán los polinomios $f(x)$ y $g(x)$.

5.1.2. Cifrado

Supongamos que Bianca desea enviar un mensaje a Aldo. Entonces debe buscar la clave pública de Aldo y representar el mensaje como un elemento del anillo R, es decir, $m(x) \in R$. Esta representación se puede lograr con algún código estándar de conversión de caracteres, por ejemplo, el código ASCII es una opción común. Luego, elige un polinomio de manera aleatoria $r(x) \in T(d, d)$ y calcula el polinomio,

$$e(x) = ph(x)r(x) + m(x) \mod q.$$

Entonces el polinomio $e(x)$ es el mensaje cifrado.

5.1.3. Descifrado

Suponemos que Aldo recibe el mensaje de Bianca y para descifrarlo realiza los siguientes cálculos:

$$a(x) = f(x)e(x) \mod q.$$

Se realiza el ω-levantamiento del polinomio $a(x)$ a R, obteniendo un polinomio $a'(x)$ y calcula,

$$b(x) = F_p(x)a'(x) \mod p.$$

Se realiza el ω-levantamiento de $b(x)$ a R, obteniendo un polinomio $b'(x)$, el cual resulta ser el mensaje original descifrado $m(x)$.

5.2. Fallo de descifrado

En el proceso de descifrado se tiene que calcular el producto de polinomios $f(x)e(x)$ módulo q y sustituyendo el polinomio $e(x)$, encontramos la siguiente expresión polinomial:

$$A \equiv f(x)e(x) \equiv pg(x)r(x) + f(x)m(x) \text{ (mód } q).$$

En este producto de polinomios se puede dar la posibilidad de que los coeficientes de este polinomio cambien cuando se aplica la reducción módulo q y eso pone en riesgo al descifrado en el sentido de que el mensaje descifrado no será igual que el original.

Cuando desciframos en NTRU tratamos de asegurar que los coeficientes del polinomio $a(x) = pg(x)r(x)+f(x)m(x) \mod q$ se encuentren en el intervalo $\left(-\frac{q}{2}, \frac{q}{2}\right]$; así, de manera similar en ETRU queremos encontrar la probabilidad de que todos los coeficientes del polinomio A se encuentren en un dominio fundamental que genere el lattice (q). Esto se debe a que ahora no tenemos un intervalo, sino una región del plano complejo (dominio fundamental) en donde queremos que pertenezcan los coeficientes del polinomio A.

El siguiente análisis es un modelo de probabilidad de fallo para el descifrado de ETRU propuesto por K. Jarvis en [20].

La forma explícita del polinomio A es:

$$\begin{aligned}
A &= pg(x)r(x) + f(x)m(x) \quad \text{mód } q \\
&= p(g_0 + g_1x + g_2x^2 + \cdots + g_{n-1}x^{n-1})(r_0 + r_1x + r_2x^2 + \cdots + r_{n-1}x^{n-1}) + \\
&\quad (f_0 + f_1x + f_2x^2 + \cdots + f_{n-1}x^{n-1})(m_0 + m_1x + m_2x^2 + \cdots + m_{n-1}x^{n-1}) \quad \text{mód } q \\
&= p\sum_{j=0}^{n-1}\sum_{k+\ell\equiv j}(g_kx^k)(r_\ell x^\ell) + \sum_{j=0}^{n-1}\sum_{k+\ell\equiv j}(f_kx^k)(m_\ell x^\ell) \\
&= p\sum_{j=0}^{n-1}\sum_{k+\ell\equiv j} g_kr_\ell x^j + \sum_{j=0}^{n-1}\sum_{k+\ell\equiv j} f_km_\ell x^k,
\end{aligned}$$

en donde $k+\ell \equiv j$ representa la congruencia $k+\ell \equiv j$ (m*ód* n). Si a los términos u_kx^k de cada polinomio los representamos simplemente por u_k, con está notación tenemos que,

$$A = p\sum_{j=0}^{n-1}\sum_{k+\ell\equiv j} g_kr_\ell + \sum_{j=0}^{n-1}\sum_{k+\ell\equiv j} f_km_\ell.$$

Notemos que cualquier número eiseniano se puede escribir en su forma cartesiana de la siguiente manera: $\alpha = a + b\omega = a + \frac{b}{2}(-1 + \sqrt{3}i) \in \mathbb{Z}[\omega]$. Además, se tiene que $Re(\alpha) = a - \frac{b}{2}$ y $Im(\alpha) = \frac{b}{2}\sqrt{3}$. De este modo podemos reescribir los términos del polinomio A como sigue:

$$\sum_{j=0}^{n-1}\sum_{k+\ell\equiv j} g_kr_\ell = \sum_{j=0}^{n-1}\sum_{k+\ell\equiv j} Re(g_kr_\ell) + \sum_{j=0}^{n-1}\sum_{k+\ell\equiv j} Im(f_km_\ell)i.$$

Aplicamos la definición de producto de números complejos,

$$(a+bi)(c+di) = (ac-bd)+(ad+bc)i,$$

para tener que

$$p\sum_{j=0}^{n-1}\sum_{k+\ell\equiv j} g_k r_\ell = \left(Re(p)\sum_{j=0}^{n-1}\sum_{k+\ell\equiv j} Re(g_k r_\ell) - Im(p)\sum_{j=0}^{n-1}\sum_{k+\ell\equiv j} Im(g_k r_\ell)\right) + \left(Re(p)\sum_{j=0}^{n-1}\sum_{k+\ell\equiv j} Im(g_k r_\ell) + Im(p)\sum_{j=0}^{n-1}\sum_{k+\ell\equiv j} Re(g_k r_\ell)\right) i.$$

Por lo tanto, el j-ésimo término del polinomio A está dado por:

$$\begin{aligned} A_j = &\left(Re(p)\sum_{k+\ell\equiv j} Re(g_k r_\ell) - Im(p)\sum_{k+\ell\equiv j} Im(g_k r_\ell) + \sum_{k+\ell\equiv j} Re(f_k m_\ell)\right) + \\ &\left(Re(p)\sum_{k+\ell\equiv j} Im(g_k r_\ell) + Im(p)\sum_{k+\ell\equiv j} Re(g_k r_\ell) + \sum_{k+\ell\equiv j} Im(f_k m_\ell)\right) i. \end{aligned} \tag{5.1}$$

Considerando el teorema del límite central, [[35], capítulo 7, teorema 7.4], podemos suponer que si el parámetro n de la sección 5.1.1 es relativamente grande, los coeficientes $Re(A_j)$ y $Im(A_j)$ tienden a una distribución normal, donde cada término $Re(A_j), Im(A_j)$ son variables aleatorias independientes.

Sea $b \in U(\mathbb{Z}[\omega])$. Para polinomios $r(x) \in T(d_r, d_r), g(x) \in T(d_g, d_g)$ que tienen sus coeficientes en el conjunto $U(\mathbb{Z}[\omega]) \cup \{0\}$, la probabilidad de que ocurran los eventos $r_\ell = b$ y $g_k = b$ es $P(r_\ell = b) = \dfrac{d_r}{n}$ y $P(g_k = b) = \dfrac{d_g}{n}$ respectivamente. Claramente los coeficientes r_ℓ, g_k están en $U(\mathbb{Z}[\omega]) \cup \{0\}$ y así tenemos que $Re(r_\ell g_k)$ toma valores en el conjunto $\{\pm 1, \pm 1/2, 0\}$. A continuación buscamos los elementos r_ℓ, g_k que satisfagan que $Re(r_\ell g_k) = 1$:

$$\begin{aligned} &r_\ell = \pm 1 \;\; y \;\; g_k = \pm 1 \;\Rightarrow\; r_\ell g_k = 1 \;\Rightarrow\; Re(r_\ell g_k) = 1, \\ &r_\ell = \pm\omega \;\; y \;\; g_k = \pm\omega^2 \;\Rightarrow\; r_\ell g_k = \omega^3 = 1 \;\Rightarrow\; Re(r_\ell g_k) = 1, \\ &r_\ell = \pm\omega^2 \;\; y \;\; g_k = \pm\omega \;\Rightarrow\; r_\ell g_k = \omega^3 = 1 \;\Rightarrow\; Re(r_\ell g_k) = 1. \end{aligned}$$

Para encontrar los elementos r_ℓ, g_k que satisfagan la condición $Re(r_\ell g_k) = -1$, basta con tomar el negativo del valor de r_ℓ en las ecuaciones anteriores y así obtenemos los resultados correspondientes a $Re(r_\ell g_k) = -1$, como se requiere en este caso.

Por lo tanto, los elementos r_ℓ, g_k tales que $Re(r_\ell g_k) = 1$ los podemos sintetizar en el conjunto,

$$\begin{aligned} A_1 &= \{(r_\ell, g_k) : Re(r_\ell g_k) = 1\} \\ &= \left\{(1,1), (-1,-1), (\omega, \omega^2), (-\omega, -\omega^2), (\omega^2, \omega), (-\omega^2, -\omega)\right\}. \end{aligned}$$

De manera análoga se define el conjunto $A_{-1} = \{(r_\ell, g_k) : Re(r_\ell g_k) = -1\}$.

Notemos que el conjunto A_1 se puede reescribir de la siguiente forma,

$$A_1 = \left\{(\alpha, \alpha^{-1}) : \alpha \in U(\mathbb{Z}[\omega])\right\}.$$

Con los resultados anteriores podemos concluir que,

$$\begin{aligned} P(Re(g_k r_\ell) = 1) &= P(g_k r_\ell = -1) \\ &= \sum_{(\alpha,\beta)\in A_1} P((g_k, r_\ell) = (\alpha, \beta)) \\ &= \sum_{\alpha \in U(\mathbb{Z}[\omega])} P(r_\ell = \alpha \text{ y } g_k = \alpha^{-1}) \\ &= \sum_{\alpha \in U(\mathbb{Z}[\omega])} P(r_\ell = \alpha) P(g_k = \alpha^{-1}) \\ &= 6\left(\frac{d_g}{n}\right)\left(\frac{d_r}{n}\right) = \frac{6 d_g d_r}{n^2}. \end{aligned}$$

De manera análoga podemos encontrar los elementos $r_\ell, g_k \in U(\mathbb{Z}[\omega])$, tales que $Re(r_\ell g_k) = \frac{1}{2}$. Recordemos que estamos considerando $Re(r_\ell g_k)$ como la parte real que resulta de expresar a los coeficientes $r_\ell g_k$ en su forma cartesiana. Entonces vemos que,

$$\begin{aligned} &r_\ell = \pm 1 \;\; y \;\; g_k = \mp\omega \;\Rightarrow\; r_\ell g_k = -\omega = \frac{1}{2} - \frac{\sqrt{3}}{2} i \;\Rightarrow\; Re(r_\ell g_k) = \frac{1}{2}, \\ &r_\ell = \pm\omega \;\; y \;\; g_k = \mp 1 \;\Rightarrow\; r_\ell g_k = -\omega = \frac{1}{2} - \frac{\sqrt{3}}{2} i \;\Rightarrow\; Re(r_\ell g_k) = \frac{1}{2}, \\ &r_\ell = \pm\omega^2 \;\; y \;\; g_k = \mp\omega^2 \;\Rightarrow\; r_\ell g_k = -\omega = \frac{1}{2} - \frac{\sqrt{3}}{2} i \;\Rightarrow\; Re(r_\ell g_k) = \frac{1}{2}, \\ &r_\ell = \pm 1 \;\; y \;\; g_k = \mp\omega^2 \;\Rightarrow\; r_\ell g_k = -\omega^2 = -\left(-\frac{1}{2} - \frac{\sqrt{3}}{2} i\right) \;\Rightarrow\; Re(r_\ell g_k) = \frac{1}{2}, \\ &r_\ell = \pm\omega \;\; y \;\; g_k = \mp\omega \;\Rightarrow\; r_\ell g_k = -\omega^2 = -\left(-\frac{1}{2} - \frac{\sqrt{3}}{2} i\right) \;\Rightarrow\; Re(r_\ell g_k) = \frac{1}{2}, \\ &r_\ell = \pm\omega^2 \;\; y \;\; g_k = \mp 1 \;\Rightarrow\; r_\ell g_k = -\omega^2 = -\left(-\frac{1}{2} - \frac{\sqrt{3}}{2} i\right) \;\Rightarrow\; Re(r_\ell g_k) = \frac{1}{2}. \end{aligned}$$

Dichos elementos r_ℓ y g_k, tales que $Re(r_\ell g_k) = \dfrac{1}{2}$ los podemos reescribir en el siguiente conjunto,

$$\begin{aligned} A_{1/2} &= \left\{(r_\ell, g_k) : Re(r_\ell g_k) = \frac{1}{2}\right\} \\ &= \left\{ \begin{array}{l} (-1,\omega), (\omega,-1), (-1,\omega^2), (\omega^2,-1), (1,-\omega), (-\omega,1) \\ (-\omega^2,1), (1,-\omega^2), (-\omega^2,\omega^2), (\omega^2,-\omega), (\omega,-\omega), (-\omega,\omega) \end{array} \right\}. \end{aligned}$$

De manera similar podemos definir el conjunto,

$$A_{-1/2} = \left\{(r_\ell, g_k) : Re(r_\ell g_k) = -\frac{1}{2}\right\}.$$

Luego, calculamos la siguiente probabilidad:

$$\begin{aligned} P\left(Re(g_k r_\ell) = \frac{1}{2}\right) &= P\left(Re(g_k r_\ell) = -\frac{1}{2}\right) \\ &= \sum_{(\alpha,\beta)\in A_{1/2}} P((g_k, r_\ell) = (\alpha,\beta)) \\ &= \sum_{(\alpha,\beta)\in A_{1/2}} P(g_k = \alpha \text{ y } r_\ell = \beta) \\ &= \sum_{(\alpha,\beta)\in A_{1/2}} P(g_k = \alpha)P(r_\ell = \beta) \\ &= 12\left(\frac{d_g}{n}\right)\left(\frac{d_r}{n}\right) = \frac{12 d_g d_r}{n^2}. \end{aligned}$$

Además, el valor esperado y varianza de la variable aleatoria $Re(g_k r_\ell)$ son,

$$E(Re(g_k r_\ell)) = 1\left(\frac{6 d_g d_r}{n^2}\right) + (-1)\left(\frac{6 d_g d_r}{n^2}\right) + \frac{1}{2}\left(\frac{12 d_g d_r}{n^2}\right) - \frac{1}{2}\left(\frac{12 d_g d_r}{n^2}\right) = 0,$$

y

$$\begin{aligned} Var(Re(g_k r_\ell)) &= E(Re(g_k r_\ell)^2) - E(Re(g_k r_\ell))^2 \\ &= E(Re(g_k r_\ell)^2) + 0 \\ &= (1)^2\left(\frac{6 d_g d_r}{n^2}\right) + (-1)^2\left(\frac{6 d_g d_r}{n^2}\right) + \left(\frac{1}{2}\right)^2\left(\frac{12 d_g d_r}{n^2}\right) + \left(\frac{-1}{2}\right)^2\left(\frac{12 d_g d_r}{n^2}\right) \\ &= \frac{18 d_g d_r}{n^2}. \end{aligned}$$

De manera similar, como $g_k r_\ell \in U(\mathbb{Z}[\omega]) \cup \{0\}$, tenemos que el término $Im(g_k r_\ell)$ toma alguno de los valores en el conjunto $\left\{\sqrt{3}/2, -\sqrt{3}/2, 0\right\}$ y con un razonamiento análogo a los casos anteriores obtenemos que,

$$P\left(Im(g_k r_\ell) = \frac{\sqrt{3}}{2}\right) = P\left(Im(g_k r_\ell) = -\frac{\sqrt{3}}{2}\right) = \frac{12 d_g d_r}{n^2}.$$

Luego, el valor esperado y varianza son:

$$E(Im(g_k r_\ell)) = \frac{\sqrt{3}}{2}\left(\frac{12 d_g d_r}{n^2}\right) - \frac{\sqrt{3}}{2}\left(\frac{12 d_g d_r}{n^2}\right) = 0.$$

$$Var(Im(g_k r_\ell)) = \left(\frac{\sqrt{3}}{2}\right)^2\left(\frac{12 d_g d_r}{n^2}\right) + \left(-\frac{\sqrt{3}}{2}\right)^2\left(\frac{12 d_g d_r}{n^2}\right) = \frac{18 d_g d_r}{n^2}.$$

De lo anterior notemos que $Var(Re(g_k r_\ell)) = Var(Im(g_k r_\ell))$. Dado que los coeficientes f_k y m_ℓ de los polinomios $f(x)$ y $m(x)$, respectivamente, satisfacen las mismas condiciones que los coeficientes g_k y r_ℓ, es fácil verificar que,

$$Var(Re(f_k m_\ell)) = Var(Im(f_k m_\ell)) = \frac{18 d_f d_m}{n^2}.$$

Consideremos dos propiedades que se satisfacen para la varianza de variables aleatorias: $Var(X+Y) = Var(X) + Var(Y)$, siempre que las variables aleatorias X, Y sean independientes y $Var(bX) = b^2X$, para $b \in \mathbb{R}$. Si suponemos que cada uno de los términos $g_k r_\ell$ y $f_k m_\ell$, para toda $k + \ell \equiv j \pmod{n}$ son independientes, entonces,

$$Var\left(\sum_{k+\ell\equiv j} Re(g_k r_\ell)\right) = Var\left(\sum_{k+\ell\equiv j} Im(g_k r_\ell)\right) = n\frac{18d_f d_m}{n^2} = \frac{18d_g d_r}{n}.$$

y

$$Var\left(\sum_{k+\ell\equiv j} Re(f_k m_\ell)\right) = Var\left(\sum_{k+\ell\equiv j} Im(f_k m_\ell)\right) = n\frac{18d_f d_m}{n^2} = \frac{18d_f d_m}{n}.$$

Calculamos la varianza a la parte real e imaginaria del j-ésimo término A_j en (5.1),

$$\begin{aligned}\sigma_R^2 &= Re(p)^2 Var\left(\sum_{k+\ell\equiv j} Re(g_k r_\ell)\right) + Im(p)^2 Var\left(\sum_{k+\ell\equiv j} Im(g_k r_\ell)\right) + Var\left(\sum_{k+\ell\equiv j} Re(f_k m_\ell)\right)\\ &= Re(p)^2\frac{18d_g d_r}{n} + Im(p)^2\frac{18d_g d_r}{n} + \frac{18d_f d_m}{n}\\ &= \frac{18}{n}(|p|^2 d_g d_r + d_f d_m),\end{aligned}$$

donde $|\cdot|$ es la norma usual de los números complejos. Por las igualdades de las varianzas de las partes reales e imaginarias en los términos $Re(g_k r_\ell), Im(g_k r_\ell), Re(f_k m_\ell), Im(f_k m_\ell)$, vistas anteriormente, podemos escribir la siguiente ecuación $\sigma_R^2 = \sigma_I^2$.

Dado $A = pg(x)r(x) + f(x)m(x) \mod q$, buscamos calcular la probabilidad de que sus coeficientes se encuentren en el dominio fundamental generado por el ideal (q). Esto es, queremos calcular la probabilidad de que los coeficientes de A se encuentren en el hexágono H (dominio fundamental) inscrito en los círculos de radio $\sqrt{\nu(q)/4}$ y $\sqrt{\nu(q)/3}$.

Supongamos que X e Y son variables aleatorias que siguen una distribución normal bivariante de la siguiente manera: sea la variable aleatoria bidimensional (x, y), con densidad conjunta,

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{x-\mu_x}{\sigma_x}\frac{y-\mu_y}{\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\},$$

donde $-\infty < x, y < \infty$, μ_x, σ_x son la media y desviación estándar de X, μ_y, σ_y son la media y desviación estándar de Y, respectivamente y ρ es el coeficiente de correlación de X e Y. Estos conceptos de estadística, así como los mencionados en la discusión anterior para calcular probabilidades, esperanzas, varianzas, etc., se pueden encontrar en un libro de texto de probabilidad y estadística estándar, por ejemplo en [35] y [26].

Con el supuesto de que $X = Re(A_j)$ y $Y = Im(A_j)$ son variables aleatorias independientes que se distribuyen de manera normal, entonces tal como se vio en el análisis anterior se tiene que $\rho = 0$ y $E(Re(g_k r_\ell)) = \mu_R = 0$, $E(Im(g_k r_\ell)) = \mu_I = 0$. Además, la desviación estándar

de estas mismas variables X e Y es $\sigma_R = \sigma_I = \sigma = \sqrt{\frac{18}{n}(|p|^2 d_g d_r + d_f d_m)}$. Al sustituir estos valores en la función $f(x, y)$ obtenemos,

$$f(x,y) = \frac{1}{2\pi\sigma\sigma\sqrt{1-0}} exp\left\{-\frac{1}{2(1-0)}\left[\left(\frac{x-0}{\sigma}\right)^2 - 2\cdot 0\frac{x-0}{\sigma}\frac{y-0}{\sigma} + \left(\frac{y-0}{\sigma}\right)^2\right]\right\}$$

$$= \frac{1}{2\pi\sigma^2} exp\left\{-\frac{x^2+y^2}{2\sigma^2}\right\}.$$

La probabilidad de que A_j esté en el dominio fundamental H es:

$$P(A_j \in H) = \iint_H f(x,y)dxdy.$$

Por lo tanto, suponiendo que los n coeficientes del polinomio A son independientes, la probabilidad de que se encuentren en el hexágono H es:

$$P\left(\bigwedge_{j=0}^{n-1} A_j \in H\right) = \left(\iint_H f(x,y)dxdy\right)^n.$$

Esto nos lleva a la siguiente conclusión. Suponiendo que los coeficientes del polinomio $A = pg(x)r(x) + f(x)m(x)$ son independientes y que las partes real e imaginaria de cada coeficiente A_j son independientes, la probabilidad de fallo del descifrado ETRU es aproximadamente,

$$1 - \left(\iint_H f(x,y)dxdy\right)^n,$$

donde H es el hexágono inscrito en los círculos de radio $\sqrt{\nu(q)/4}$ y $\sqrt{\nu(q)/3}$.

5.3. Ataque de lattices al cifrado ETRU

Actualmente se conocen algunos ataques comunes al cifrado ETRU, algunos de ellos se desarrollan en [20]. En esta sección analizamos el ataque de lattices al cifrado ETRU.

Buscamos una analogía al ataque visto en la sección 3.4 del capítulo 3, en donde la matriz que genera al lattice L^{NT} es,

$$B^{NT} = \begin{bmatrix} \lambda I & H \\ 0 & qI \end{bmatrix},$$

donde I es una matriz identidad de tamaño $n \times n$, H es la matriz circulante de la clave pública $h(x)$ y $\lambda \in \mathbb{R}$ es su respectiva constante de equilibrio.

En la sección 4.1.1 del capítulo 4 vimos la manera de representar un número eiseniano mediante una matriz y al producto de dos números complejos de forma matricial en la base $\mathcal{B} = \{1, i\}$, es decir, si $z = x + yi \in \mathbb{Z}[\omega]$, definimos,

$$[T]_{\mathcal{B}} = \begin{bmatrix} c & -d \\ d & c \end{bmatrix},$$

donde $T : \mathbb{C} \to \mathbb{C}$ es la transformación lineal dada por $T(z) = z\beta$, y $\beta = c + d\omega$ es fijo. Entonces realizamos cálculos análogos en $\mathbb{Z}[\omega]$ de la siguiente manera.

Si $\alpha = a + b\omega \in \mathbb{Z}[\omega]$, definimos,

$$\langle \alpha \rangle = \begin{bmatrix} a & b \\ -b & a-b \end{bmatrix}, \tag{5.2}$$

y vemos que si $\beta = c + d\omega$ es otro número eiseniano, el producto matricial,

$$[c, d] \begin{bmatrix} a & b \\ -b & a-b \end{bmatrix} = [ac - bd, bc + ad - bd] \tag{5.3}$$

es la representación del producto de los enteros eisenianos α y β en la base $\{1, \omega\}$.

Para cada matriz A de tamaño $n \times n$, con entradas en $\mathbb{Z}[\omega]$, establecemos $\langle A \rangle$, que es una matriz de tamaño $2n \times 2n$, con entradas a_{ij} y en donde los a_{ij} son matrices de tamaño 2×2 dadas por $\langle a_{ij} \rangle$ y consideremos la siguiente proposición.

Proposición 5.3.1. *Denotaremos al conjunto de matrices cuadradas con entradas en $\mathbb{R}$ por $Mat_{n\times n}$. Definimos la función $\psi : \mathbb{Z}[\omega] \to Mat_{2\times 2}$, dada por,*

$$\psi(a + b\omega) = \begin{bmatrix} a & b \\ -b & a-b \end{bmatrix}.$$

Entonces ψ es un homomorfismo inyectivo.

Recordemos que la ecuación (5.2) consiste en expresar a un número eiseniano en forma matricial, es decir,

$$\langle a + b\omega \rangle = \begin{bmatrix} a & b \\ -b & a-b \end{bmatrix}.$$

Esta notación permite reescribir a la función ψ como sigue,

$$\psi(a + b\omega) = \langle a + b\omega \rangle .$$

Demostración. Sean $z_1 = a + b\omega, z_2 = c + d\omega$ elementos de $\mathbb{Z}[\omega]$ cualesquiera, entonces,

$$\begin{aligned} \psi(z_1 z_2) = \psi((a + b\omega)(c + d\omega)) &= \psi((ac - bd) + (ad + bc - bd)\omega) \\ &= \begin{bmatrix} ac - bd & ad + bc - bd \\ -ad - bc + bd & ac - ad - bc \end{bmatrix} \\ &= \langle (ac - bd) + (ad + bc - bd)\omega \rangle . \end{aligned}$$

Por otro lado, vemos que,

$$\psi(z_1)\psi(z_2) = \psi(a + b\omega)\psi(c + d\omega) = \langle a + b\omega \rangle \langle c + d\omega \rangle = \begin{bmatrix} a & b \\ -b & a-b \end{bmatrix} \begin{bmatrix} c & d \\ -d & c-d \end{bmatrix} =$$

$$= \begin{bmatrix} ac - bd & ad + b(c-d) \\ -bc + (-d)(a-b) & -bd + (a-c)(c-d) \end{bmatrix} = \begin{bmatrix} ac - bd & ad + bc - bd \\ -bc - ad + bd & ac - ad - bc \end{bmatrix}.$$

Por lo tanto, se cumple que $\psi(z_1 z_2) = \psi(z_1)\psi(z_2)$, es decir, ψ es un homomorfismo.

Ahora, supongamos que $\psi(z_1) = \psi(z_2)$, entonces se sigue que,

$$\begin{gathered}\langle z_1\rangle = \langle z_2\rangle \\ \begin{bmatrix} a & b \\ -b & a-b \end{bmatrix} = \begin{bmatrix} c & d \\ -d & c-d \end{bmatrix},\end{gathered}$$

es decir, se satisface el sistema de ecuaciones,

$$\begin{aligned} a &= b \\ b &= d \\ -b &= -d \\ a-b &= c-d. \end{aligned}$$

Entonces concluimos que $a = c$ y $b = d$; y así, vemos que $a + b\omega = c + d\omega$ o que $z_1 = z_2$. Por lo tanto, ψ es un homomorfismo inyectivo. □

Una utilidad de la proposición anterior es que nos permite realizar adecuadamente multiplicaciones matriciales de la forma $\langle z\rangle D$, donde D es una matriz de tamaño 2×2, y z es un número eiseniano.

Ejemplo 5.3.1. *Si $q = a + b\omega \in \mathbb{Z}[\omega]$, el producto de $\langle q\rangle$ con una matriz identidad de tamaño 2×2 se escribe como sigue:*

$$\langle q\rangle I_{2\times 2} = \begin{bmatrix} a & b \\ -b & a-b \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a & b \\ -b & a-b \end{bmatrix} = \langle q\rangle .$$

Sea B^{ET} una matriz dada por,

$$B^{ET} = \begin{bmatrix} \lambda\langle I\rangle & \langle H\rangle \\ 0 & \langle qI\rangle \end{bmatrix},$$

en donde $\lambda\langle I\rangle$ es una matriz identidad de tamaño $n \times n$ de bloques de la forma $\lambda\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\langle H\rangle$ es la matriz circulante por bloques $\langle h_i\rangle$ de la clave pública $h(x)$ y $\langle qI\rangle$ denota a la matriz de tamaño $n \times n$ de bloques de la forma $\langle q\rangle I_{2\times 2}$.

Verificaremos que la matriz B^{ET} es una base para el lattice que se genera a partir de la clave pública $h(x)$ del criptosistema ETRU, al cual denotaremos por L^{ET}.

Sean $\alpha_1, \alpha_2, \ldots, \alpha_{2n}$ enteros eisenianos, donde cada α_i es de la forma $\alpha_i = c_i + d_i\omega$. Consideramos el vector asociado a la concatenación repetida de cada uno de estos números eisenianos de la siguiente manera,

$$\bar{\alpha}_i = [(c_i, d_i), (c_i, d_i), \ldots, (c_i, d_i)]_{1\times 4n}.$$

Los vectores columna de la matriz B^{ET} son de la forma,

$$C_1 = \begin{bmatrix} \lambda I_{2\times 2} \\ \langle 0\rangle \\ \vdots \\ \langle 0\rangle \end{bmatrix}_{4n\times 2}, C_2 = \begin{bmatrix} \langle 0\rangle \\ \lambda I_{2\times 2} \\ \vdots \\ \langle 0\rangle \end{bmatrix}_{4n\times 2}, \ldots, C_{2n} = \begin{bmatrix} \langle h_{n-1}\rangle \\ \langle h_{n-2}\rangle \\ \vdots \\ \langle q\rangle I_{2\times 2} \end{bmatrix}_{4n\times 2}$$

Para verificar la independencia lineal de las columnas de B^{ET}, tenemos lo siguiente:

Supongamos que,

$$0 = \bar{\alpha_1}C_1 + \bar{\alpha_2}C_2 + \cdots + \bar{\alpha}_{2n}C_{2n}, \tag{5.4}$$

donde $0 = 0 + 0\omega$.

Observemos, por ejemplo, que el primer sumando $\bar{\alpha_1}C_1$ de la ecuación (5.4) tiene la forma:

$$\begin{aligned}
\bar{\alpha_1}C_1 &= [(c_1, d_1), (c_1, d_1), \ldots, (c_1, d_1)]_{1\times 4n} \begin{bmatrix} \lambda I_{2\times 2} \\ \langle 0 \rangle \\ \vdots \\ \langle 0 \rangle \end{bmatrix}_{4n\times 2} \\
&= \lambda[c_1, d_1]I_{2\times 2} + \sum_{i=2}^{2n}[c_1, d_1]\langle 0 \rangle \\
&= \lambda[c_1, d_1]\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + [0, 0] \\
&= \lambda[c_1, d_1] + [0, 0] = \lambda[c_1, d_1].
\end{aligned}$$

De manera similar es fácil ver que en el resto de los productos matriciales $\bar{\alpha}_i C_i$, para cada $i = 2, 3, \ldots, 2n$ se obtendrá un vector de la forma $[c_i x, d_i y]$ con x, y números reales distintos de cero. Entonces, cada factor $\bar{\alpha}_i C_i$ es un vector cero siempre que el vector $[c_i, d_i] = [0, 0]$, se sigue que los enteros eisenianos α_i son cero, es decir, se tiene que $\alpha_i = c_i + d_i\omega = 0 + 0\omega$.

Por lo tanto, la ecuación (5.4) se satisface siempre que los coeficientes eisenianos α_i son cero, para $i = 1, 2, \ldots, 2n$. Luego, las columnas $C_1, C_2, \ldots, C_{2n}$ de la matriz B^{ET} son linealmente independientes.

Vemos que la dimensión de L^{ET} es $4n$, pues en B^{ET} se pueden expresar $4n$ columnas linealmente independientes. Por lo tanto, el lattice L^{ET} contiene un conjunto finito de vectores linealmente independientes, se sigue que $C_1, C_2, \ldots, C_{2n}$ forman una base para este lattice.

Notemos que el vector $[f', u] \in \mathbb{Z}^{4n}$ representa los coeficientes de los polinomios $f'(x)$ y $u(x)$. Ahora, cada coeficiente de f', $f_i = a_i + b_i\omega$, es representado en su forma vectorial por $f_i = [a_i, b_i]$ y de forma similar para el vector u. Notemos que esto a su vez nos dice que $f', u \in \mathbb{Z}^{2n}$. Entonces tenemos que,

$$\begin{aligned}
[f'\ u]B^{ET} &= [f'\ u]\begin{bmatrix} \lambda\langle I \rangle & \langle H \rangle \\ 0 & \langle qI \rangle \end{bmatrix} \\
&= [f'\lambda\langle I \rangle, f'\langle H \rangle + u\langle qI \rangle] \\
&= [\lambda f'\langle I \rangle, f'\langle H \rangle + u\langle qI \rangle].
\end{aligned}$$

Así, obtenemos que,

$$[f'\ u]B^{ET} = [\lambda f', f'\langle H \rangle + u\langle qI \rangle]. \tag{5.5}$$

Veamos de manera más explícita al producto de matrices anterior; notemos que la matriz $\langle H \rangle$ tiene la forma,

$$\langle H \rangle = \begin{bmatrix} \langle h_0 \rangle & \langle h_1 \rangle & \cdots & \langle h_{n-1} \rangle \\ \langle h_{n-1} \rangle & \langle h_0 \rangle & \cdots & \langle h_{n-2} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle h_1 \rangle & \langle h_{n-2} \rangle & \cdots & \langle h_0 \rangle \end{bmatrix},$$

y si representamos a las columnas de la matriz $\langle H \rangle$ por $\langle H \rangle = [H_0, H_1, \ldots, H_{n-1}]$, entonces el producto $f' \langle H \rangle$ se expresa como sigue:

$$f' \langle H \rangle = ([a_0, b_0], [a_1, b_1], \ldots, [a_{n-1}, b_{n-1}])(H_0, H_1, \ldots, H_{n-1}). \tag{5.6}$$

Tomamos el primer producto $f'H_0$ de $f' \langle H \rangle$ y vemos que tiene la siguiente forma:

$$f'H_0 = ([a_0, b_0], [a_1, b_1], \ldots, [a_{n-1}, b_{n-1}]) \begin{bmatrix} \langle h_0 \rangle \\ \langle h_{n-1} \rangle \\ \vdots \\ \langle h_1 \rangle \end{bmatrix} = [a_0, b_0] \langle h_0 \rangle + \cdots + [a_{n-1}, b_{n-1}] \langle h_1 \rangle . \tag{5.7}$$

Como los productos $[a_i, b_i] \langle h_j \rangle$ representan el producto de coeficientes "números eisenianos", vistos de manera matricial, y como la clave pública es de la forma $h(x) = F_q(x)g(x) \mod q$, entonces cada producto $[a_i, b_i] \langle h_j \rangle$ es congruente con algún coeficiente $g_k = c_k + d_k\omega$ (número eiseniano) de $g(x)$ módulo q, que en notación vectorial se expresa $g_k = [c_k, d_k]$, por lo que,

$$[a_i, b_i] \langle h_j \rangle \equiv [c_k, d_k] \text{ (mód } q).$$

Entonces al tomar la reducción módulo q en la suma (5.7) se tiene que,

$$f'H_0 = [a_0, b_0] \langle h_0 \rangle + \cdots + [a_{n-1}, b_{n-1}] \langle h_1 \rangle = [c_r, d_r] \text{ (m\'od } q),$$

donde $[c_r, d_r]$ es la representación vectorial del eiseniano $g_r = c_r + d_r\omega$.

Por lo tanto, $f'H_j = [c_r, d_r] \mod q$, para cada j, k con $0 \leq j < n$ y $0 \leq k < n$. Repitiendo el proceso anterior para cada j, podemos concluir que el vector en (5.6) es congruente con el vector g' módulo q, de modo que podemos escribir $f' \langle H \rangle = g' \mod q$.

Entonces si consideramos el factor $f' \langle H \rangle + qu \mod q$ de la ecuación (5.5), se sigue que,

$$[f' \ u]B^{ET} = [\lambda f', f' \langle H \rangle + u \langle qI \rangle] = [\lambda f', g'].$$

Además, notemos que el vector $[\lambda f', g']$ está contenido en el lattice L^{ET}. Este vector es relativamente corto, por lo que, podemos usar técnicas de reducción de bases como lo es el algoritmo LLL para intentar recuperar el vector $[\lambda f', g']$, es decir, recuperar la clave privada del cifrado ETRU, $(f(x), g(x))$.

Sea $\alpha = a + b\omega \in \mathbb{Z}[\omega]$. Definimos la siguiente función norma sobre el anillo $\mathbb{Z}[\omega]$, $N : \mathbb{Z}[\omega] \to \mathbb{R}$, dada por $N(\alpha) = \sqrt{a^2 + b^2}$, a dicha función la llamamos la función norma euclidiana o simplemente norma euclidiana de $\mathbb{Z}[\omega]$. En el apéndice B se proporciona una prueba que justifica que la función N es una norma para $\mathbb{Z}[\omega]$.

Observación 5.3.1. *Notemos que si un número eiseniano $z = a+b\omega$ está dentro de una región fundamental H, generada por el ideal (q), con $q = c + d\omega$, se tiene que $a \leq c$ y $b \leq d$, y con esto se sigue que $N(z) \leq N(q)$.*

Observación 5.3.2. *Un inconveniente en la implementación del algoritmo LLL es que éste encuentra vectores en relación con la norma euclidiana, la cual difiere de la norma usual en $\mathbb{Z}[\omega]$. Si $\alpha = a+b\omega \in \mathbb{Z}[\omega]$, entonces este elemento tiene norma euclidiana $N(\alpha) = \sqrt{a^2 + b^2}$, mientras que la norma algebraica o norma usual en $\mathbb{Z}[\omega]$ para α es $||\alpha|| = \sqrt{\nu(\alpha)} = \sqrt{a^2 - ab + b^2}$. Entonces, para implementar este ataque se considerará la norma euclidiana de $\mathbb{Z}[\omega]$ ya que dicha norma mejora el funcionamiento del algoritmo LLL, además de que es similar a la norma euclidiana de los números complejos $\mathbb{C}$.*

En particular, para resaltar la importancia de qué norma de $\mathbb{Z}[\omega]$ se va a considerar en el siguiente análisis vemos que las unidades $\pm\omega^2$ tienen norma euclidiana igual a $\sqrt{2}$, que es mayor que su norma algebraica, la cual es igual a 1. Este simple cálculo puede generar cambios significativos para un mismo análisis considerando una u otra norma en dicho análisis.

Recordemos que la elección del polinomio $f(x)$ se realiza en el conjunto $T(d+1,d)$, es decir, que $f(x)$ tiene $d+1$ coeficientes iguales a 1 , d coeficientes de cada una del resto de las unidades de $\mathbb{Z}[\omega]$, y el resto de sus coeficientes iguales a cero, además de que el polinomio $x-1$ no sea un factor de $f(x)$.

Entonces, $f(x)$ puede tomar como forma más general la siguiente:

$$\begin{aligned} f(x) = {} & 1 + x + \cdots + x^d + (-1)x^{d+1} + (-1)x^{d+2} + \cdots + (-1)x^{2d} + \\ & + \omega x^{2d+1} + \omega x^{2d+2} + \cdots + \omega x^{3d} + (-\omega)x^{3d+1} + (-\omega)x^{3d+2} + \cdots + (-\omega)x^{4d} + \\ & + \omega^2 x^{4d+1} + \omega^2 x^{4d+2} + \cdots + \omega^2 x^{5d} + (-\omega)^2 x^{5d+1} + (-\omega)^2 x^{5d+2} + \cdots + (-\omega)^2 x^{6d}. \end{aligned} \tag{5.8}$$

Análogamente como la elección de $g(x)$ se realiza en el conjunto $T(d,d)$, la manera más general que puede tomar el polinomio $g(x)$ es:

$$\begin{aligned} g(x) = {} & 1 + x + \cdots + x^{d-1} + (-1)x^d + (-1)x^{d+1} + \cdots + (-1)x^{2d-1} + \\ & + \omega x^{2d} + \omega x^{2d+1} + \cdots + \omega x^{3d-1} + (-\omega)x^{3d} + (-\omega)x^{3d+1} + \cdots + (-\omega)x^{4d-1} + \\ & + \omega^2 x^{4d} + \omega^2 x^{4d+1} + \cdots + \omega^2 x^{5d-1} + (-\omega)^2 x^{5d} + (-\omega)^2 x^{5d+1} + \cdots + (-\omega)^2 x^{6d-1}. \end{aligned}$$

En el apéndice B se proporciona una prueba de que la función $\bar{N} : \mathbb{Z}[\omega]^n \to \mathbb{R}$, dada por $\bar{N}(z_1, z_2, \ldots, z_n) = \sqrt{N(z_1)^2 + N(z_1)^2 + \cdots + N(z_n)^2}$ es una norma para el conjunto $\mathbb{Z}[\omega]^n$, a la que llamamos norma euclidiana en $\mathbb{Z}[\omega]^n$.

Note que la norma euclidiana en $\mathbb{Z}[\omega]^n$ se ha expresado con la notación $\bar{N}$, mientras que la norma euclidiana para $\mathbb{Z}[\omega]$ se ha expresado con la notación N. En la práctica expresaremos simplemente por N a cualquiera de estas dos normas cuando no hay una posible confusión entre las dos normas y se especificará por $\bar{N}$ o N cuando sea necesario.

Entonces, podemos calcular la norma euclidiana de los polinomios $f(x)$ y $g(x)$ como sigue: $N(f(x)) := N(f_0 + f_1 x + \cdots + f_{n-1}x^{n-1}) = N((f_0, f_1, \ldots, f_{n-1}))$, luego, tenemos que,

$$\begin{aligned} N(f(x))^2 = {} & \underbrace{N(1)^2 + \cdots + N(1)^2}_{d+1 \text{ veces}} + \underbrace{N(-1)^2 + \cdots + N(-1)^2}_{d \text{ veces}} + \underbrace{N(\omega)^2 + \cdots + N(\omega)^2}_{d \text{ veces}} + \\ & + \underbrace{N(-\omega)^2 + \cdots + N(-\omega)^2}_{d \text{ veces}} + \underbrace{N(\omega^2)^2 + \cdots + N(\omega^2)^2}_{d \text{ veces}} + \underbrace{N(-\omega^2)^2 + \cdots + N(-\omega^2)^2}_{d \text{ veces}} \\ = {} & (d+1)N(1)^2 + dN(-1)^2 + dN(\omega)^2 + dN(-\omega)^2 + dN(\omega^2)^2 + dN(-\omega^2)^2 \\ = {} & (d+1)(1)^2 + d(1)^2 + d(1)^2 + d(1)^2 + d(\sqrt{2})^2 + d(\sqrt{2})^2 \\ = {} & 8d + 1. \end{aligned}$$

Notemos que en la penúltima igualdad anterior se han calculado y sustituido las normas euclidianas de las unidades de $\mathbb{Z}[\omega]$, las cuales son,

$$\begin{aligned} N(\pm 1) &= \sqrt{(\pm 1)^2 + 0} = 1, \\ N(\pm\omega) &= \sqrt{0^2 + (\pm 1)^2} = 1, \\ N(\pm\omega^2) = N(\pm(-1-\omega)) &= \sqrt{(\mp 1)^2 + (\mp 1)^2} = \sqrt{2}. \end{aligned}$$

Por lo tanto, $N(f(x)) = \sqrt{8d+1}$.

De manera similar podemos realizar los mismos cálculos anteriores para obtener la norma euclidiana del polinomio $g(x)$, la cual resulta ser $N(g(x)) = \sqrt{8d}$.

Ahora bien, si τ representa la longitud del vector objetivo $(\lambda f, g)$, entonces,

$$\tau = \sqrt{\lambda^2 N(f(x))^2 + N(g(x))^2}.$$

Al sustituir los valores de $N(f(x))$ y $N(g(x))$ en τ, obtenemos que,

$$\tau = \sqrt{\lambda^2[8d+1] + 8d}.$$

Dado que en este caso se usó la norma euclidiana para calcular a τ, usaremos la notación $\tau_c = \tau$ para referirnos a este cálculo de τ con la norma euclidiana de $\mathbb{Z}$ y $\mathbb{Z}[\omega]^n$.

De acuerdo a la observación 5.3.2 podemos realizar el cálculo de τ, pero ahora con respecto a la norma usual de $\mathbb{Z}[\omega]$, al cual denotaremos por τ_e, es decir, $\tau_e = \tau$.

Ahora bien, consideramos otra función norma para el conjunto $\mathbb{Z}[\omega]^n$, la cual se define de la siguiente manera $||\ || : \mathbb{Z}[\omega]^n \to \mathbb{R}$ dada por,

$$||(z_1, z_2, \ldots, z_n)|| = \sqrt{||z_1||^2 + ||z_2||^2 + \cdots + ||z_n||^2}.$$

A esta norma la llamaremos norma usual en $\mathbb{Z}[\omega]^n$. Entonces con esta norma podemos volver a calcular el valor de τ, al cual denotamos por τ_e.

Primero calculamos la norma de $f(x)$ y $g(x)$,

$$\begin{aligned}
||f(x)||^2 &= \underbrace{||1||^2 + \cdots + ||1||^2}_{d+1 \text{ veces}} + \underbrace{||-1||^2 + \cdots + ||-1||^2}_{d \text{ veces}} + \underbrace{||\omega||^2 + \cdots + ||\omega||^2}_{d \text{ veces}} + \\
&+ \underbrace{||-\omega||^2 + \cdots + ||-\omega||^2}_{d \text{ veces}} + \underbrace{||\omega^2||^2 + \cdots + ||\omega^2||^2}_{d \text{ veces}} + \underbrace{||-\omega^2||^2 + \cdots + ||-\omega^2||^2}_{d \text{ veces}} \\
&= (d+1)||1||^2 + d||-1||^2 + d||\omega||^2 + d||-\omega||^2 + d||\omega^2||^2 + d||-\omega^2||^2 \\
&= (d+1)(1)^2 + d(1)^2 + d(1)^2 + d(1)^2 + d(1)^2 + d(1)^2 \\
&= 6d+1,
\end{aligned}$$

por lo que, $||f(x)|| = \sqrt{6d+1}$.

$$\begin{aligned}
||g(x)||^2 &= \underbrace{||1||^2 + \cdots + ||1||^2}_{d \text{ veces}} + \underbrace{||-1||^2 + \cdots + ||-1||^2}_{d \text{ veces}} + \underbrace{||\omega||^2 + \cdots + ||\omega||^2}_{d \text{ veces}} + \\
&+ \underbrace{||-\omega||^2 + \cdots + ||-\omega||^2}_{d \text{ veces}} + \underbrace{||\omega^2||^2 + \cdots + ||\omega^2||^2}_{d \text{ veces}} + \underbrace{||-\omega^2||^2 + \cdots + ||-\omega^2||^2}_{d \text{ veces}} \\
&= (d)||1||^2 + d||-1||^2 + d||\omega||^2 + d||-\omega||^2 + d||\omega^2||^2 + d||-\omega^2||^2 \\
&= (d)(1)^2 + d(1)^2 + d(1)^2 + d(1)^2 + d(1)^2 + d(1)^2 \\
&= 6d,
\end{aligned}$$

por lo que, $||g(x)|| = \sqrt{6d}$.

De este modo concluimos que τ_e queda determinado de la siguiente manera,

$$\tau_e = \sqrt{\lambda^2[6d+1] + 6d}.$$

Observemos que el valor de τ calculado con la norma euclidiana es $\tau_c = \sqrt{\lambda^2[8d+1] + 8d}$, que es mayor al calculado con la norma usual de $\mathbb{Z}[\omega]$, $\tau_e = \sqrt{\lambda^2[6d+1] + 6d}$. En la práctica uno desearía que el valor de τ_c fuese más pequeño que el valor de τ_e, sin embargo se requiere usar τ_c para lograr un funcionamiento adecuado con el algoritmo LLL.

El vector $(\lambda f, g)$ es un vector relativamente corto en el lattice L^{ET}, de modo que se puede emplear el algoritmo LLL para encontrar un vector de longitud τ, y así, dicho vector puede resultar ser algún vector corto o alguna rotación de $(\lambda f, g)$.

5.3.1. Elección de la constante de equilibrio λ

Análogamente a la sección 3.4.1 del capítulo 3, buscamos que la constante λ sea elegida para maximizar la eficiencia de encontrar vectores cortos en el lattice L^{ET}. Entonces, se desea maximizar la razón s/τ, donde s es la longitud esperada del vector distinto de cero más corto en L^{ET} y τ la longitud objetivo, es decir, del vector $(\lambda f, g)$ en L^{ET}.

Si τ representa la longitud objetivo de $(\lambda f, g)$ respecto a la norma euclidiana, entonces,

$$\tau = \sqrt{\lambda^2 N(f(x))^2 + N(g(x))^2},$$

donde $N(f(x))$ es la norma euclidiana de la representación vectorial de $f(x)$, y similarmente para $g(x)$.

La heurística gaussiana vista en la sección 2.5 del capítulo 2 establece que la longitud esperada del vector distinto de cero más corto es,

$$s = \sqrt{\frac{d}{2\pi e}}(det(L))^{1/d},$$

donde d es la dimensión del lattice L.

Calculamos el determinante del lattice L^{ET},

$$det(L^{ET}) = det(B^{ET}) = det\left(\begin{bmatrix} \lambda\langle I\rangle & \langle H\rangle \\ 0 & \langle qI\rangle \end{bmatrix}\right) =$$

$$det\left(\left[\begin{array}{cccc|cccc} \langle\lambda\rangle & 0 & \cdots & 0 & \langle h_0\rangle & \langle h_1\rangle & \cdots & \langle h_{n-1}\rangle \\ 0 & \langle\lambda\rangle & \cdots & 0 & \langle h_0\rangle & \langle h_1\rangle & \cdots & \langle h_{n-1}\rangle \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \langle\lambda\rangle & \langle h_0\rangle & \langle h_1\rangle & \cdots & \langle h_{n-1}\rangle \\ \hline 0 & 0 & \cdots & 0 & \langle q\rangle & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \langle q\rangle & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & \langle q\rangle \end{array}\right]\right) = det(\mathbf{B}),$$

luego, como la penúltima matriz anterior es una matriz triangular superior por bloques, se sigue

que su determinante es igual a,

$$\begin{aligned} det(L^{ET}) = det(\mathbf{B}) &= \prod_{i=1}^{n} det\left(\langle\lambda\rangle\right) \prod_{i=1}^{n} \left(\langle q\rangle\right) \\ &= \left[\lambda det\left(\langle\lambda\rangle\right)\right]^n \left[det\left(\langle q\rangle\right)\right]^n \\ &= \left[det\left(\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right)\right]^n \left[det\left(\begin{bmatrix} a & b \\ -b & a-b \end{bmatrix}\right)\right]^n \\ &= \lambda^{2n}\nu(q)^n. \end{aligned}$$

En la última y la penúltima igualdad anterior se consideró el ejemplo 5.3.1, para obtener el cálculo de $det(L^{ET})$.

Por otro lado, vimos anteriormente que la dimensión del lattice L^{ET} es $4n$. Entonces, tenemos que la longitud esperada del vector más corto es,

$$s = \sqrt{\frac{4n}{2\pi e}}\left(\lambda^{2n}(\nu(q)^{2n}\right)^{1/4n} = \sqrt{\frac{2n}{\pi e}}\lambda^{1/2}\nu(q)^{1/2} = \sqrt{\frac{2n\lambda||q||_2}{\pi e}}.$$

La relación s/τ queda de la siguiente manera,

$$\begin{aligned} s/\tau &= \sqrt{\frac{\frac{2n\lambda||q||_2}{\pi e}}{\lambda^2[8d+1]+8d}} \\ &= \sqrt{\frac{2n\lambda||q||_2}{\pi e\lambda^2(8d+1)+\pi e(8d)}} \\ &= \sqrt{\frac{2n\lambda||q||_2}{\pi e(\lambda^2 N(f(x))^2 + N(g(x))^2)}}, \end{aligned}$$

así que maximizar la relación s/τ es lo mismo que maximizar el término,

$$\frac{\lambda}{\lambda^2 N(f(x))^2 + N(g(x))^2}.$$

Observe que esta es la misma expresión que encontramos en la sección 3.4.1 del capítulo 3, por lo que podemos elegir $\lambda = N(g(x))/N(f(x))$, y con los supuestos de que $f(x), g(x) \in T(d+1, d)$, se puede considerar que $N(g(x)) \approx N(f(x))$ y así podemos tomar $\lambda = 1$.

Note que dependiendo de nuestra elección del parámetro d de la sección 5.1.1, el valor $\tau = \sqrt{\lambda^2 N(f(x))^2 + N(g(x))^2}$ puede no ser la longitud del vector real más corto en L^{ET}. Sin embargo, con argumentos similares a la proposición 3.4.1 del capítulo 3, se tiene que el vector $[\underbrace{(\lambda,0),\ldots,(\lambda,0)}_{n\ veces},\underbrace{(0,0),\ldots,(0,0)}_{n\ veces}]$ que tiene norma euclidiana $\sqrt{n\lambda^2}$, que está en el lattice L^{ET}. Como en el caso de NTRU esto no parece afectar nuestro éxito en encontrar un vector objetivo, siempre y cuando se consideren parámetros adecuados como los de la sección 5.1.1.

Capítulo 6

Firma digital con NTRU

En este capítulo se darán los conceptos básicos de función hash, y la definición de esquema de firma digital, necesaria para implementar un esquema de firma digital con los criptosistemas NTRU y ETRU. Algunos conceptos de este capítulo son estándar y los podemos encontrar en la literatura, por ejemplo, en [10] y [11]. También, presentaremos el esquema de firma digital con NTRU propuesto por J. Hoffstein et al. en [7], junto con un análisis de seguridad para los ataques de lattices y ataques de transcripciones que se emplean en dicho esquema de firma digital.

6.1. Funciones hash

El objetivo de las funciones **hash** es detectar modificaciones de los mensajes transmitidos que estén cifrados o no. Este objetivo a menudo se denomina "integridad de datos". Una función criptográfica hash puede garantizar la integridad de los datos en determinados entornos. Se usa una función hash para construir una pequeña "huella dactilar"de algunos datos.

Se debe suponer que la huella dactilar se almacena en un lugar seguro. Luego, incluso si los datos se almacenan en un lugar inseguro, su integridad se puede verificar volviendo a calcular la huella dactilar y comprobando que esta no haya cambiado. Si h es una función hash y x representa algunos datos, por ejemplo, x podría ser una cadena binaria de longitud arbitraria. Entonces, una huella dactilar correspondiente se define como $y = h(x)$. Esta huella dactilar también se le conoce como resumen del mensaje. El resumen del mensaje suele ser una cadena binaria corta; 160 bits o 256 bits son opciones comunes.

Además, podemos suponer que y se almacena en un lugar seguro, pero no x. Si por ejemplo se cambia x a x_0, entonces esperamos que el resumen del mensaje antiguo "y" no sea también un resumen de mensaje para x_0, pero si este es realmente el caso, entonces el hecho de que x ha sido alterado puede detectarse simplemente calculando el resumen del mensaje $y_0 = h(x_0)$ y verificando que $y_0 \neq y$.

Las funciones hash tienen múltiples usos en la criptografía, entre ellos la generación de sucesiones pseudoaleatorias y el uso de esquemas de firma digital, que usaremos más adelante.

También se tienen funciones hash h_k que dependen de alguna clave k. Una función hash con clave se utiliza a menudo como "código de autenticación de mensajes" o MAC, por sus siglas en inglés. Suponga que los usuarios Alice y Bob comparten una clave secreta, k, que determina una función hash, digamos d_k.

Para un mensaje x, el apéndice de autenticación correspondiente o simplemente apéndice, es $y = h_k(x)$. Alice y Bob pueden calcular este apéndice. El par (x, y) se puede transmitir a través de un canal inseguro de Alice a Bob (o de Bob a Alice). Supongamos que Bob recibe el par (x, y) de Alice. Luego, puede verificar si $y = h_k(x)$ volviendo a calcular el apéndice. Si esta condición se cumple, entonces Bob está seguro de que ni x ni y fueron alterados por un adversario, siempre que la familia hash sea "segura". En particular, a Bob se le asegura que el mensaje x se origina en Alice.

En cuanto a la diferencia de integridad, observe la distinción entre la garantía de integridad de los datos proporcionados por una función hash sin clave, en contraposición a una función hash con clave. En el caso de una función hash sin clave, el resumen del mensaje debe almacenarse de forma segura para que un adversario no pueda modificarlo.

Por otro lado, si Alice y Bob utilizan una clave secreta k para especificar la función hash que están usando, entonces pueden transmitir tanto los datos como el apéndice de autenticación sobre un canal inseguro.

Definición 6.1.1. *Una **familia hash** es una tupla de cuatro conjuntos (X, Y, K, H), donde se cumplen las siguientes condiciones:*

1. *X es un conjunto de mensajes posibles (usualmente X es el conjunto de todas las posibles cadenas binarias finitas).*
2. *Y es un conjunto finito de posibles resúmenes de mensajes o apéndices de autenticación.*
3. *K, el espacio de claves, es un conjunto finito de claves posibles.*
4. *Para cada $k \in K$, hay una función hash $h_k \in H$ tal que $h_k : X \to Y$.*

En la definición anterior X puede ser finito o infinito, Y es siempre un conjunto finito. Para crear una firma digital se utilizan únicamente las funciones hash sin clave, por lo que podemos suponer que el conjunto K solo tiene un elemento.

6.2. Definición de firma digital

Se utiliza una firma manuscrita "convencional" adjunta a un documento para especificar la persona responsable del mismo. La firma se utiliza en situaciones cotidianas como escribir una carta, retirar dinero de un banco, firmar un contrato, etc.

Un esquema de firma es un método para firmar un mensaje almacenado en forma electrónica. Una firma digital no se adjunta físicamente al mensaje que está firmando. Se sugiere que siempre se debe evitar que se reutilice un mensaje digital firmado, por ejemplo en casos de cheques o pagarés.

6.2.1. Componentes de una firma digital

Un esquema de firma digital consta de dos componentes: un algoritmo de firma y un algoritmo de verificación.

Alice puede firmar un mensaje x utilizando un algoritmo de firma (privado), sig_k, que depende de una clave privada k. La firma $sig_k(x)$ resultante se puede verificar posteriormente

utilizando un algoritmo de verificación público ver_k.

Dado un par (x, y), donde x es el mensaje e y es una supuesta firma en x, el algoritmo de verificación devuelve una respuesta de válida o no válida, dependiendo de si y es o no una firma válida para el mensaje x.

Definición 6.2.1. *Un **esquema de firma** es una tupla de cinco conjuntos (P, A, K, S, V), donde se cumplen las siguientes condiciones:*

1. *P es un conjunto de mensajes posibles.*
2. *A es un conjunto finito de posibles firmas, a menudo son cadenas binarias.*
3. *K, el espacio de claves, es un conjunto finito de claves posibles.*
4. *Para cada $k \in K$, hay un algoritmo de firma $sig_k \in S$ y un algoritmo de verificación correspondiente $ver_k \in V$.*

Cada $sig_k : P \to A$ y $ver_k : P \times A \to \{válida, novália\}$ son funciones tales que la siguiente ecuación se satisface. Para cada mensaje $x \in P$ y para cada firma $y \in A$

$$ver_k(x, y) = \begin{cases} válida, & si \ \ y = sig_k(x) \\ no\ válida, & si \ \ y \neq sig_k(x). \end{cases}$$

Un par (x, y), con $x \in P$ e $y \in A$ se denomina mensaje firmado.

En cuanto a la eficiencia e información privada en algunos esquemas de firma, el algoritmo de firmas funciona como un algoritmo aleatorio. Para cada clave $k \in K$, las funciones sig_k y ver_k deben ser funciones de tiempo polinomial. El algoritmo, ver_k, será público y el algoritmo de firma sig_k será privado.

De este modo, dado un mensaje x, debería ser computacionalmente inviable para cualquier otra persona que no sea Alice calcular una firma tal que $ver_k(x, y) = válida$. Notemos que puede haber más de una y para una x dada, dependiendo de cómo se defina la función ver_k. Si un tercer usuario, digamos Oscar, puede calcular un par (x, y) tal que $ver_k(x, y) = válida$ y x no fue previamente firmado por Alice, entonces a la firma y se llama falsificación generada por el adversario Oscar. De manera informal, una firma falsificada es una firma válida producida por alguien que no sea Alice.

6.3. Esquema de firma digital para el criptosistema RSA

Para ilustrar el funcionamiento de un esquema de firma digital presentamos el esquema de firma RSA, el cual se basa en lo siguiente:

Criptosistema: Esquema de firma digital RSA (versión básica).

Consideremos números enteros n, p, q, e, d y defina,

$$K = \{(n, p, q, e, d) : n = pq, \ p, q \text{ son primos distintos}, ed \equiv 1 \ (\text{mód } \varphi(n))\},$$

donde $\varphi(n)$ es la función φ de Euler.

Los valores n y e son la clave pública y los valores p, q y d son la clave privada. Tomamos $P = A = \mathbb{Z}_n$, y para $K = (n, p, q, e, d)$, defina $sig_k(x) = x^d \mod n$ y $ver_k(x, y) = válida$ si y solo si $x \equiv y^e \pmod n$, para $x, y \in \mathbb{Z}_n$.

Alice es la única persona que puede crear la firma por que el parámetro d es privado. El algoritmo de verificación también utiliza la regla de encriptación RSA. Observe que cualquiera puede verificar una firma porque "e" es público, esto se logra eligiendo una y aleatoria y calculando $x = y^e \mod n$; entonces $sig_k(x) = (y^e)^d = y$ es una firma válida para el mensaje x.

6.4. Requisitos de seguridad para esquemas de firmas

En esta sección analizamos lo que significa que un esquema de firmas sea seguro.

Como fue el caso con un criptosistema, necesitamos especificar un modelo de ataque, el objetivo del adversario y el tipo de seguridad proporcionada por el esquema. Recordemos que el modelo de ataque está definido por la información disponible para el adversario y en el caso de esquema de firmas, se suelen considerar los siguientes tipos de ataque:

1. **Ataque con solo la clave pública.**
   Oscar posee la clave pública de Alice, es decir la función de verificación, ver_k.

2. **Ataque con mensaje conocido o ataque de transcripciones.**
   Oscar posee una lista de mensajes previamente firmados por Alice, digamos las parejas $(x_1, y_1), (x_2, y_2), \ldots$, donde las x_i son mensajes y las y_i son firmas de Alice en estos mensajes, es decir, $y_i = sig_k(x_i),\ i = 1, 2, \ldots$

3. **Ataque con mensaje elegido.**
   Oscar solicita las firmas de Alice en una lista de mensajes. Por lo tanto, elige los mensajes $x_1, x_2, \ldots$, y Alice proporciona sus firmas en estos mensajes, a saber, $y_i = sig_k(x_i),\ i = 1, 2, \ldots$

6.4.1. Objetivos del adversario

Consideraremos varios posibles objetivos del adversario:

1. **Ruptura total.**
   Oscar es capaz de determinar la clave privada de Alice, es decir, la función de firma sig_k. Por lo tanto, puede crear firmas válidas en cualquier mensaje.

2. **Falsificación selectiva.**
   Con alguna probabilidad no despreciable, Oscar puede crear una firma válida en un mensaje elegido por otra persona. En otras palabras si Oscar recibe un mensaje x, entonces puede determinar (con cierta probabilidad) una firma y, tal que $ver_k(x, y) = válida$. El mensaje x no debe ser uno que haya sido firmado previamente por Alice.

3. **Falsificación existencial.**
   Oscar puede crear una firma válida para al menos un mensaje. En otras palabras, Oscar puede crear un par (x, y), donde x es un mensaje y $ver_k(x, y) = válida$. El mensaje x no debe ser uno que haya sido firmado previamente por Alice.

El esquema de firmas no puede ser incondicionalmente seguro, ya que Oscar puede probar todas las firmas posibles $y \in A$, para un mensaje x dado, utilizando el algoritmo público ver_k,

hasta que encuentre una firma válida. Entonces si se le da el tiempo suficiente, Oscar siempre puede falsificar la firma de Alice en cualquier mensaje.

Por lo tanto, tal como fue el caso de los criptosistemas de clave pública, nuestro objetivo es encontrar esquemas de firmas que sean seguros desde el punto de vista informático o demostrable.

6.4.2. Ejemplos de ataques

Ya sabemos que Oscar puede construir un mensaje firmado válido con el criptosistema RSA, eligiendo una firma y y luego calculando x, tal que $ver_k(x, y) = \textit{válida}$. Esto sería una falsificación existencial usando un ataque con solo la clave pública.

Otro tipo de ataque a la firma digital con RSA se basa en la propiedad multiplicativa de RSA. Suponga que $y_1 = sig_k(x_1)$ y $y_2 = sig_k(x_2)$ son dos mensajes cualesquiera que fueron firmados previamente por Alice. Entonces $ver_k(x_1x_2 \mod n, y_1y_2 \mod n) = \textit{válida}$, y por lo tanto Oscar puede crear la firma válida $y_1y_2 \mod n$ para el mensaje $x_1x_2 \mod n$. Este es un ejemplo de falsificación existencial mediante un ataque con mensaje conocido.

Otro ataque posible es el siguiente. Supongamos que Oscar quiere falsificar una firma en el mensaje x, donde x fue posiblemente elegido por otra persona. Para él es muy sencillo encontrar $x_1, x_2 \in \mathbb{Z}_n$, tales que $x = x_1x_2 \mod n$. Ahora, Oscar le pide a Alice su firma en los mensajes x_1 y x_2, que denotamos por y_1 e y_2 respectivamente. Entonces como en el ataque anterior y_1y_2 mód n es una firma del mensaje $x = x_1x_2 \mod n$. Esta es una firma de falsificación selectiva, usando un ataque con mensaje elegido.

6.4.3. Firmas y funciones hash

Actualmente los esquemas de firmas casi siempre se usan junto con una función hash criptográfica segura (pública).

Las funciones hash $h : \{0,1\}^* \to P$, donde $\{0,1\}^*$ es el conjunto de cadenas binarias finitas, tomará un mensaje de longitud arbitraria y producirá un resumen de mensaje (hash) de un tamaño especifico, por ejemplo 224 bits es una opción popular.

El resumen del mensaje se firmará usando un esquema de firma (P, A, K, S, V). El uso de una función hash y su respectiva firma se representa en forma de diagrama en la siguiente figura.

$$\begin{array}{ccccc} \text{mensaje} & & \text{huella del mensaje} & & \text{firma} \\ x & \to & z = h(x) & \to & y = sig_k(z) \\ x \in \{0,1\}^* & & z \in Z & & y \in Y \end{array}$$

Figura 6.1: Esquema de firma digital con función hash.

Suponga que Alice quiere firmar un mensaje x, que es una cadena de bits de longitud arbitraria. Primero construye el resumen del mensaje $z = h(x)$, y luego calcula la firma en z, es decir $y = sig_k(z)$. Luego, transmite el par ordenado (x, y) por algún canal de transmisión.

Ahora la verificación puede ser realizada, por cualquier persona, reconstruyendo primero el resumen del mensaje $z = h(x)$, usando la función hash pública, y luego, verificando que $ver_k(z, y) = \textit{válida}$. Debemos tener cuidado de que el uso de una función hash no debilite la

seguridad del esquema de firma, ya que es el resumen del mensaje el que está firmado, no el mensaje.

Será necesario que la función hash en cuestión satisfaga determinadas propiedades para prevenir diversos ataques. Algunas de las propiedades que deseamos que satisfaga una función hash para ser resistente a los ataques ya mencionados se explican a continuación.

1. Preimagen

Problema: **Preimagen**.
Instancia: Una función hash $h : X \to Y$ y un elemento $y \in Y$.
Encontrar: $x \in X$, tal que $h(x) = y$.

Dado un resumen de mensaje y, el problema **Preimagen** pregunta si se puede encontrar un elemento $x \in X$, tal que $h(x) = y$. Tal valor sería una preimagen de y.

Si **Preimagen** se puede resolver, para algún $y \in Y$ dado, entonces el par (x, y) es un par válido. A menudo se dice que una función hash para la que **Preimagen** no puede resolverse de manera eficiente es unidireccional *resistente a preimagen*.

2. Segunda Preimagen

Problema: **Segunda Preimagen**.
Instancia: Una función hash $h : X \to Y$ y un elemento $x \in X$.
Encontrar: $x_0 \in X$, tal que $x_0 \neq x$ y $h(x_0) = h(x)$.

Dado un mensaje x, el problema **Segunda Preimagen** pregunta si se puede encontrar un elemento $x_0 \neq x$, de manera que $h(x_0) = h(x)$. Comenzamos con x, que es una preimagen de y, y buscamos hasta encontrar un valor x_0 que sería una segunda preimagen de y. Si esto se puede hacer, entonces $(x_0, h(x_0))$ es un par válido. A menudo se dice que una función hash para la que la **Segunda Preimagen** no se puede resolver de manera eficiente es *resistente a segunda preimagen*.

3. Colisión

Problema: **Colisión**.
Instancia: Una función hash.
Encontrar: $x_0, x \in X$, tal que $x_0 \neq x$ y $h(x_0) = h(x)$.

El problema **Colisión** pregunta si se puede encontrar algún par de entradas (x_0, x), tales que $h(x_0) = h(x)$. Como era de esperar, esto se llama colisión. Una solución a este problema produce dos pares válidos (x, y) y (x_0, y_0), donde $y = h(x) = h(x_0)$.

Hay varios escenarios en los que queremos que surja tal situación. A menudo se dice que una función hash para la que **Colisión** no se puede resolver de manera eficiente es *resistente a colisiones*.

6.4.4. Ataques

El tipo de ataque más obvio es que el adversario Oscar empiece con un mensaje firmado válido (x, y), donde $y = sig_k(h(x))$. Luego, calcula $h(x)$ e intenta encontrar $x' \neq x$, tal que

$h(x') = h(x)$.

Si Oscar puede hacer esto, (x', y) sería un mensaje firmado válido, entonces podemos decir que y es una firma falsificada para el mensaje x'. Se trata de una falsificación existencial que utiliza un ataque con mensaje conocido. Para prevenir este tipo de ataques requerimos que h sea resistente a segundas preimagenes.

Otro posible ataque es el siguiente:

Oscar primero encuentra dos mensajes $x \neq x'$, tales que $h(x) = h(x')$. Luego, Oscar le da x a Alice y la convence de que firme el resumen del mensaje $h(x)$, obteniendo y.

Entonces (x', y) es un mensaje firmado válido, con y una firma falsificada para el mensaje x'. Esto es una falsificación existencial que utiliza un ataque con mensaje elegido, el cual se puede prevenir si h es resistente a colisiones.

6.5. Esquema de firma digital con NTRU

Los autores Oded Goldreich, Shafi Goldwasser y Shai Halevi, en 1995, diseñaron un esquema de firma digital (GGH) basado en lattices, y que tiene su seguridad en resolver el problema del vector más corto (CVP) en un lattice [8]. Las firmas GGH forman la base de algunos esquemas de firma digital basados en lattices, por ejemplo para el esquema de firma digital NTRU. Por otro lado los autores del criptosistema NTRU, Joseph H. Silverman, Jeffery Hoffstein, Jill Pipher, han propuesto esquemas de firma digital con NTRU, sin embargo las primeras versiones han resultado poco resistentes a algunos ataques, por ejemplo, a los ataques de transcripción. Dichos ataques propuestos por diferentes autores motivaron el diseño de una versión de firma digital con NTRU propuesta por H. Silverman et al. en [11] que fuera resistente a ataques de lattices, transcripciones, etc. En esta sección estudiaremos el esquema de firma digital con NTRU, que evita la vulnerabilidad de los ataques de transcripción y que analiza cómo es que un atacante podría generar firmas falsas mediante ataques de lattices.

6.5.1. Muestreo por rechazo y seguridad de transcripciones

Recordemos que los esquemas de firma digital difieren de los criptosistemas de clave pública en que cada documento con su respectiva firma, (x, y), revela potencialmente información sobre la clave privada, como es el caso de los esquemas de firma digital GGH y NTRU. Un ataque de transcripción es un método para intentar recuperar la clave privada a través de una larga lista de posibles firmas:

$$(x_1, y_1), (x_2, y_2), \ldots, (x_t, y_t). \tag{6.1}$$

Esto es una vulnerabilidad a la que la mayoría de los esquemas de firma quedan expuestos. En versiones iniciales de esquema de firma digital con lattices, por ejemplo el esquema de firma digital GGH o alguna de las primeras versiones de firma digital con NTRU, cada firma revela un vector en el lattice que genera el esquema de firma digital GGH, (L), o con NTRU, (L^{NT}), de la forma:

$$y = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n.$$

El atacante no conoce los a_i ni los v_i, pero tomando un promedio ponderado aproximado sobre una transcripción de la forma (6.1), puede construir una imagen del dominio fundamental

de NTRU o GGH, es decir, en los lattices L^{NT} o L respectivamente, que tiene la siguiente forma:

$$\{t_1v_1 + t_2v_2 + \cdots + t_nv_n : 0 \leq t_i < 1\}.$$

Usando esta imagen, el atacante puede falsificar firmas.

Algunos autores como C. Gentry, M. Szydlo et. al. en [17], propusieron un ataque potencial al demostrar que la fuga de información mediante firmas válidas reduce el problema de recuperación de claves, al de distinguir formas cuadráticas integrales. Se propuso un método heurístico para resolver este último problema, considerando un promedio de transcripción para recuperar casi de manera directa el polinomio $f(x)$. Posteriormente Phong Q. Nguyen y Oded Regev, en [30], diseñaron un algoritmo eficiente para recuperar la clave secreta en forma de un dominio fundamental a partir de un pequeño número de firmas.

V. Lyubashevsky en [21], describió cómo usar el muestreo de rechazo para eliminar por completo los ataques de transcripción en ciertos esquemas de firma basados en lattices, tales conceptos son los siguientes:

1. Incluir aleatoriedad en cada firma.
2. Rechaza las firmas "malas" y solo usa firmas "buenas".
3. Se considera de manera correcta la distribución de probabilidad del conjunto de buenas firmas, la cual será la misma para todas las claves privadas.

J. Hoffstein et al., en [11], han propuesto una versión de firma digital que es segura ante los ataques de transcripción y además permite encontrar una base para el lattices en que se lleva a cabo la firma digital, y que también es factible para el estudio de posibles ataques mediante lattices.

6.5.2. Algoritmos para generar algunos polinomios

En esta sección se estudian algunos algoritmos para encontrar polinomios con propiedades particulares, por ejemplo algoritmos para poder generar polinomios aleatorios con distribución uniforme, algoritmos para convertir polinomios aleatorios en cadenas de bytes y viceversa. Estos algoritmos se pueden encontrar en [6].

Algoritmo 1. Algoritmo generador de polinomios ternarios de peso fijo con $(d+1)$ coeficientes 1 y (d) coeficientes -1.
Este algoritmo se inicia con los parámetros n, d y una cadena de bits s para la aleatoriedad.

Entrada: n, d, s.

Salida: Un polinomio ternario aleatorio de grado $n-1$.

Operaciones: El algoritmo se calcula mediante la siguiente sucesión de pasos o una equivalente:

1. Establecer $f(x) := 0$.
2. Estableces $t := 0$.
3. Mientras $t \leq d$, hacer:

   3.1 $s_1 =$ extraer una cadena de $(log\ n + 1)$ bits de la cadena s.

3.2 $k =$ escribir el entero en binario que representa s_1.

3.3 $i = k \mod n$.

3.4 Si $f_i = 0$,

I) Establecer $f_i := 1$.

II) Establecer $t := t + 1$.

4. Establecer $t := 0$.

5. Mientras $t < d$, hacer:

5.1 $s_1 =$ extraer $(log\ n + 1)$ bits de la cadena s.

5.2 $k =$ escribir el entero en binario que representa s_1.

5.3 $i = k \mod n$.

5.4 Si $f_i = 0$,

I) Establecer $f_i := -1$.

II) Establecer $t := t + 1$.

6. Salida: $f(x)$.

Algoritmo 2. Algoritmo generador de polinomios mód p uniforme.
Se crea una instancia de un generador de polinomios uniformes con los parámetros n, p y se considera una cadena de bits s para la aleatoriedad.
Produce un polinomio de grado $n - 1$ cuyos coeficientes se distribuyen de manera aleatoria uniforme entre 0 y $p - 1$.

Entrada: n, p y s.

Salida: Un polinomio ternario pseudoaleatorio $f(x)$ cuyos coeficientes se encuentran entre 0 y $p - 1$.

Operaciones: El polinomio se calculará mediante la siguiente sucesión de pasos o una equivalente:

1. Establecer $f(x) := 0$.

2. Para $(i = 0, i < n)$, hacer:

2.1 $b = log\ p + 1$.

2.2 $s = s_0 s_1 \ldots s_{b-1} s_b s_{b+1} \ldots s_{2b-1} \ldots$.

2.3 $t_i = s_{bi} s_{bi+1} \ldots s_{bi+b-1}$.

2.4 $k =$ escribir el entero en binario que representa t_i.

2.5 Establecer $f_i = k \mod p$.

2.6 Incrementar $i = i + 1$.

3. Salida: $f(x)$.

Los siguientes algoritmos sirven para convertir elementos de un anillo de polinomios a cadenas de octetos utilizando el número mínimo de bits por coeficiente.

Observación 6.5.1. *Si bien la representación de cadenas de octetos puede ser más conveniente para la aritmética de elementos de anillos de polinomios en un microprocesador, los elementos del anillo pueden almacenarse y transmitirse de manera más compacta como cadenas de bits.*

Si $(x_1, x_2, \ldots, x_n)$ es un argumento para una función hash, denotaremos por $bLen$ a la longitud de octetos tras aplicar la función hash a dicho argumento.

Algoritmo 3. Algoritmo que convierte un número entero no negativo a una cadena de bits de una longitud especifica. Lo denotaremos por (I2BSP), por sus siglas en inglés.

Entrada: Un entero no negativo x; $bLen$, longitud prevista de la cadena de bits resultante.

Salida: Una cadena de bits, B, de longitud $bLen$.

Operaciones: La cadena de bits se calculará mediante los siguientes pasos o una sucesión de pasos equivalente:

1. Si $x \geq 2^{xLen}$, genera, "entero demasiado grande" y detener el algoritmo.
2. Escriba el entero x en su representación única $xLen - bit$ en base 2:

$$x = x_{xLen-1} \times 2^{xLen-1} + x_{xLen-2} \times 2^{xLen-2} + \cdots + x_1 \times 2 \times x_0,$$

   donde $x_i = 0$ o 1 (note que uno o más bits iniciales serán cero si x es menor que 2^{xLen-1}).
3. Salida: La cadena de bits $x_{xLen-1}x_{xLen-2}\ldots x_1x_0$.

Algoritmo 4. Algoritmo que convierte un elemento de un anillo de polinomios en una cadena de bits y lo denotaremos por (RE2BSP), por sus siglas en inglés.

Entrada: Un polinomio $f(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1}$, n : la dimensión del anillo, q : módulo mayor, es decir que todos los coeficientes de $f(x)$ están entre 0 y $q - 1$.

Salida: Una cadena de bits, B.

Operaciones: La conversión se calculará mediante la siguientes sucesión de pasos o una equivalente:

1. Para $(j = 0, j < n)$, iguale A_j a la representación positiva más pequeña de $a_j \mod q$.
2. Establecer $B_j = I2BSP(A_j, bLen)$. Si alguna de las llamadas a I2BSP genera un error, emita ese error y detenga el algoritmo.
3. Salida: La cadena de bits $B = B_0B_1 \ldots B_{n-1}$.

Algoritmo 5. Algoritmo que convierte una cadena de bits a una cadena de octetos con relleno a la derecha, lo denotaremos por (BS2ROSP), por sus siglas en inglés.

Entrada: Una cadena de bits B a convertir; $oLen$: longitud prevista de la cadena de octetos resultante.

Salida: Una cadena de octetos, O correspondiente de longitud $oLen$.

Operaciones: La salida se calculará mediante los siguientes pasos o una sucesión de pasos equivalente:

1. Iguale $bLen$ a la longitud de B en bits.
2. Si $bLen > 8oLen$, emite "entrada demasiado largaz parar.

3. Agregar $(8 \cdot oLen - bLen)$ bits cero al final de B.

4. Sean $b_0 b_1 \dots b_{xLen-2} b_{xLen-1}$ los bits de B del primero al último. Para $0 \leq i < oLen - 1$, sea el octeto $O_i = b_{8i} b_{8i+1} \dots b_{8i+7}$.

5. Salida: $O = O_0 O_1 \dots O_{oLen-1}$.

Algoritmo 6. Algoritmo que convierte un elemento de un anillo de polinomios en una cadena de octetos y lo denotaremos por (RE2OSP), por sus siglas en inglés.

Entrada: Un polinomio $f(x) = a_0 + a_1 x + \cdots + a_{n-1} x^{n-1}$, n : la dimensión del anillo de polinomios, q : el módulo más grande, es decir que todos los coeficientes de $f(x)$ están entre 0 y $q - 1$; $oLen$: longitud prevista para la cadena de octetos resultante.

Salida: Una cadena de octetos correspondiente O.

Operaciones: La salida se calculará mediante la siguiente sucesión de pasos o una equivalente:

1. Convierta el polinomio $f(x)$ en una cadena de bits, bA, usando el algoritmo RE2BSP.

2. Convierta la cadena de bits bA y longitud $oLen$, en una cadena de octetos O usando el algoritmo BS2ROSP.

3. Salida: La cadena de octetos O.

Algoritmo 7. Algoritmo que convierte una cadena de octetos a una cadena de bits de una longitud específica, lo denotaremos por (ROS2BSP), por sus siglas en inglés.

Entrada: Una cadena de octetos O a convertir; $bLen$: longitud prevista de la cadena de bits resultante.

Salida: Una cadena de bits correspondiente B de longitud $bLen$.

Operaciones: La salida se calculará mediante los siguientes pasos o una sucesión de pasos equivalente:

1. Iguale $oLen$ a la longitud de O en octetos.

2. Si $bLen > 8 \cdot oLen$, emite "entrada demasiado larga" y detenga el algoritmo.

3. Para $0 \leq i < oLen - 1$, considere el octeto O_i como los bits $b_{8i} b_{8i+1} \dots b_{8i+7}$.

4. Si alguno de los bits $b_{bLen-1} \dots b_{8 \cdot oLen-1}$ no es cero, emite "bits distintos de cero encontrados después del final de la cadena de bits" y detenga el algoritmo.

5. Salida: Una cadena de bits $B = b_0 b_1 \dots b_{bLen-1}$.

Algoritmo 8. Algoritmo que convierte una cadena de bits en un entero no negativo, lo denotaremos por (BS2IP), por sus siglas en inglés.

Entrada: Una cadena de bits B a convertir ($bLen$ se usa para indicar la longitud de B).

Salida: Un entero no negativo, x, correspondiente.

Operaciones: La salida se calculará mediante los siguientes pasos o una sucesión de pasos equivalente:

1. Si B es de longitud 0, la salida es 0.
2. Sean $b_{bLen-1}b_{bLen-2}\dots b_1b_0$ los bits de B de izquierda a derecha.
3. Sea $x = b_{bLen-1}\cdot 2^{bLne-1} + b_{bLen-2}\cdot 2^{bLen-2} + \cdots + b_1\cdot 2 + b_0$.
4. Salida: x.

Algoritmo 9. Algoritmo que convierte una cadena de bits en un elemento de un anillo de polinomios, el cual denotaremos por (BS2REP), por sus siglas en inglés.

Entrada: Una cadena de bits B a convertir; n : la dimensión del anillo de polinomios, q : el módulo más grande, es decir que todos los coeficientes de $f(x)$ están entre 0 y $q-1$.

Salida: Un elemento del anillo de polinomios $f(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1}$ resultante.

Operaciones: La salida se calculará mediante los siguientes pasos o una sucesión de pasos o una equivalente:

1. Si la longitud de B no es igual a $n\cdot\lceil log_2\ q\rceil$, emite "longitud incorrecta de la cadena de bits" y detener el algoritmo.
2. Considere B como la serie de cadenas de bits $B = B_0b_1\dots B_{n-1}$, donde cada B_j tiene una longitud de $\cdot\lceil log_2\ q\rceil$ bits.
3. Para $(j = 0, j < n)$, establezca $a_j = BS2IP\ (B_j)$. Si BS2IP emite un error, emite "error".
4. Salida: $f(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1}$.

Algoritmo 10. Algoritmo que convierte una cadena de octetos O, en un polinomio $f(x)$ de un anillo de polinomios. Denotaremos este algoritmo por (OS2REP), por sus siglas en inglés.

Entrada: Una cadena de octetos O a convertir; n : la dimensión del anillo de polinomios, q : el módulo más grande, es decir que todos los coeficientes de $f(x)$ están entre 0 y $q-1$.

Salida: Un elemento de un anillo de polinomios $f(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1}$.

Operaciones: La salida se calculará mediante los siguientes pasos o una sucesión de pasos equivalente:

1. Si la longitud de la cadena O no es igual a $n\cdot\lceil log_{256}\ q\rceil$, emite "longitud incorrecta de la cadena de octetos" y detener el algoritmo.
2. Considere la cadena de octetos O en la cadena de bits bA, usando el algoritmo ROS2BSP.
3. Convierta la cadena de bits bA en un polinomio $f(x)$ usando el algoritmo BS2REP.
4. Salida: $f(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1}$.

En el siguiente algoritmo consideraremos que los coeficientes de un polinomio están en el conjunto de polinomios ternarios de $\mathbb{Z}[\omega]$.

Algoritmo 11. Algoritmo generador de polinomios ternarios de peso fijo con $(d+1)1's$ y (d) unidades de cada una del restos de las unidades de $\mathbb{Z}[\omega]$.
Este algoritmo se inicia con los parámetros n, d, ω, las unidades de los enteros de Eisenstein, $U(\mathbb{Z}[\omega]) = \left\{\pm 1, \pm\omega, \pm\omega^2\right\}$ y una cadena de bits s para la aleatoriedad.

Entrada: $n, d, s, U(\mathbb{Z}[\omega])$.

Salida: Un polinomio ternario aleatorio de grado $n-1$.

Operaciones: El algoritmo se calcula mediante la siguiente sucesión de pasos o una equivalente:

Para $\alpha \in U(\mathbb{Z}[\omega])$,

1. Establecer $f(x) := 0$.
2. Estableces $t := 0$.
3. Mientras $t < d$, hacer:
   - 3.1 $s_1 =$ extraer una cadena de $(log\ n + 1)$ bits de la cadena s.
   - 3.2 $k =$ escribir el entero en binario que representa s_1.
   - 3.3 $i = k \mod n$.
   - 3.4 Si $f_i = 0$,
     - I) Establecer $f_i := \alpha$.
     - II) Establecer $t := t + 1$.
4. Establecer $t := d + 1$.
   - 4.1 $s_1 =$ extraer una cadena de $(log\ n + 1)$ bits de la cadena s.
   - 4.2 $k =$ escribir el entero en binario que representa s_1.
   - 4.3 $i = k \mod n$.
   - 4.4 Si $f_i = 0$,
     - I) Establecer $f_i := 1$.
   - 4.5 **Salida:** $f(x)$.

En el algoritmo anterior se debe considerar que la sucesión aleatoria s no repita en cada elección de $\alpha \in \mathbb{Z}[\omega]$, por ejemplo se pueden realizar corrimientos de la sucesión s en cada elección α.

Algoritmo 12. Algoritmo generador de polinomios $\mod p$ uniforme sobre el conjunto $\mathbb{Z}[\omega]$.

Se crea una instancia de un generador de polinomios uniformes con los parámetros n, p y se considera una cadena de bits s para la aleatoriedad.

Produce un polinomio de grado $n-1$ cuyos coeficientes se distribuyen de manera aleatoria uniforme módulo p.

Entrada: $n, p = 2 + \omega$ y s.

Salida: Un polinomio ternario pseudoaleatorio $f(x)$ cuyos coeficientes se encuentran reducidos módulo p.

Operaciones: El polinomio se calculará mediante la siguiente sucesión de pasos o una equivalente:

1. Establecer $f(x) := 0$.

2. Para $(i = 0, i < n)$, hacer:

   2.1 $b = \lfloor log\ N(p) \rfloor + 1$.

   2.2 $s = s_0 s_1 \dots s_{b-1} s_b s_{b+1} \dots s_{2b-1} \dots$.

   2.3 $t_i = s_{bi} s_{bi+1} \dots s_{bi+b-1}$.

   2.4 $k =$ escribir el entero en binario que representa t_i.

   2.5 Establecer $a_i = k$.

   2.6 Incrementar $i = i + 1$.

3. Para $(j = 0, j < n)$, hacer:

   3.1 $b = \lfloor log\ N(p) \rfloor + 1$.

   3.2 $s = s_0 s_1 \dots s_{b-1} s_b s_{b+1} \dots s_{2b-1} \dots$.

   3.3 $t_j = s_{bj} s_{bj+1} \dots s_{bj+b-1}$.

   3.4 $\ell =$ escribir el entero en binario que representa t_j.

   3.5 Establecer $b_j = \ell$.

   3.6 Incrementar $j = j + 1$.

4. Establecer $f_t = k + \ell\omega$.

5. Establecer $f_t = k + \ell\omega$ módulo p.

6. Salida: $f(x)$.

En el algoritmo anterior consideramos que la sucesión s, dada en el paso 3.2 está determinada como sigue:
De la sucesión $s = s_0 s_1 \dots s_{b-1} s_b s_{b+1} \dots s_{2b-1} \dots$ utilizada en el paso 2.2 consideramos el último bit que se usó, digamos s_k, luego tomamos los bits $s_{k+1} s_{k+2} s_{k+3}$ y sumamos $s_0 + s_{k+1}, s_1 + s_{k+2}, s_2 + s_{k+3}$. Estos bits aleatorios se pueden utilizar en la primera iteración del paso 3 de este algoritmo, así, sucesivamente estos últimos tres bits los podemos sumar con los siguientes tres bits $s_{k+4} s_{k+5} s_{k+6}$, obteniendo así nuevos bits que se puede usar el la siguiente iteración, sucesivamente repetimos este proceso hasta tener suficientes bits que se utilizarán como cadena s en el paso 3.2.

6.5.3. Un poco de notación

Enseguida se presentan algunos conceptos que se utilizarán más adelante.

Considere la siguiente notación: Dado un polinomio $a(x) = a_0 + a_1 x + \cdots + a_n x^n$, con coeficientes en $\mathbb{R}$, la ***norma infinito*** de $a(x)$ es,

$$||a(x)||_\infty = ||(a_0, a_1, \dots, a_n)||_\infty = \max_i |a_i|,$$

y la ***norma euclidiana*** de $a(x)$ es,

$$||a(x)||_2 = ||(a_0, a_1, \dots, a_n)||_2 = \sqrt{a_0^2 + a_1^2 + \cdots + a_n^2}.$$

Similarmente definimos la ***norma euclidiana de dos o más polinomios concatenados*** $a(x) = a_0 + a_1 x + \cdots + a_n x^n$ y $b(x) = b_0 + b_1 x + \cdots + b_n x^n$, con coeficientes en $\mathbb{R}$, por:

$$||a(x), b(x)||_2 = ||(a_0, a_1, \dots, a_n, b_0, b_1, \dots, b_n)||_2 = \sqrt{\sum_{i=0}^{n} a_i^2 + \sum_{i=0}^{n} b_i^2}.$$

En analogía con el criptosistema NTRU, para el estudio del esquema de firma digital con NTRU, utilizaremos uno de los anillos de polinomios mencionados en la sección 3.1.1 del capítulo 3, pero con la reducción del ideal generado por el polinomio $x^n + 1$; es decir, nuestro anillo base será el anillo de polinomios $R_q = \mathbb{Z}_q[x]/\langle x^n + 1\rangle$, teniendo así la notación $R = R_q$ como anillo base, además consideramos la definición de polinomios ternarios de la sección ya mencionada en el capítulo 3.

Con la notación anterior establecemos el subconjunto $R(k)$ de R como sigue,

$$R(k) = \{f(x) \in R : ||f(x)||_\infty \leq k\},$$

donde k es un número real positivo.

Por ejemplo, $R(3/2)$ es el conjunto de polinomios ternarios.

Una matriz de rotación cíclica de un polinomio $f(x)$ sobre el anillo R es una matriz $M = (f_1, f_2, \ldots, f_n)^T$, con $f_i = f(x)x^{i-1} \mod (x^n + 1)$. Particularmente para referirnos a la norma infinito de elementos de M (vectores), establecemos la notación $NORMF$, así que $NORMF$ se calcula como $||t(x)||_\infty$, donde $t(x) = \sum_{i=0}^{n-1} f(x)x^i \mod (x^n + 1)$.

Un mensaje $m(x) \in R$ lo denotaremos por $\mu(x)$ o $\mu_i(x)$ cuando es un mensaje específico.

También, consideraremos subconjuntos de lattices L_h, que consisten en vectores de norma acotada y los denotaremos por,

$$L_h(k_1, k_2) = L_h \cap (R(k_1) \times R(k_2)),$$

donde L_h es un lattice, k_1 y k_2 son números reales positivos.

Dado un anillo R y $p \in R$, el conjunto,

$$pR = \{px : x \in R\},$$

lo llamaremos el ***conjunto de los múltiplos del elemento*** $p \in R$.

6.5.4. Definición de esquema de firma digital con NTRU

El siguiente esquema de firma digital con NTRU está basado en el esquema de firma digital propuesto por J. Hoffstein et al. en [11], que a diferencia del esquema estándar *pqNTRUSign*, también propuesto por J. Hoffstein et al. en [6] realiza un algoritmo de firma y verificación distinto, pero que sí cumple con todos los requisitos de seguridad del esquema *pqNTRUSign*, aplicando métodos de aleatoriedad, muestreo de rechazo, etc. Esto se logra distintas etapas de la generación de claves, algoritmo de firma y verificación.

Al esquema de firma digital propuesto en [11] expresado con nuestra notación lo llamaremos esquema de firma $NTRUSign$.

Definimos el esquema de firma $NTRUSign$ por la tupla:

$$(R, R \times R, pR(3/2) \times R(p/2), S, V),$$

donde

a) $R \times R$ es el producto cartesiano del anillo R consigo mismo.

b) $pR(3/2) \times R(p/2)$ es el producto cartesiano del conjunto de polinomios $pR(3/2)$, con el conjunto de polinomios $R(p/2)$.

c) S es el conjunto de algoritmos de firma $fir_k : R \to R \times R$.

d) V es el conjunto de algoritmos de verificación $ver_k : R \times R \times R \to \{válida,\ no\ válida\}$.

Note que el algoritmo de firma entrega como firma digital a un par de polinomios $(m(x), t(x))$, para el mensaje $\mu(x)$; mientras que el algoritmo de verificación tiene por entrada a la terna $(m(x), t(x), \mu(x))$. El hecho de que el algoritmo de verificación tenga como entrada a un terna de polinomios en vez de un par de polinomios se debe a que puede haber distintos pares de polinomios, digamos $(m'(x), t'(x))$ que sean firma válida para $\mu(x)$.

El esquema de firma $NTRUSign$ se lleva a cabo de la siguiente manera:

6.5.5. Generación de claves

El usuario interesado en firmar un mensaje o documento electrónico, primeramente deberá fijar los parámetros $(n, k, m, p = 2, q, d)$ y B_k, con las siguientes condiciones: n, k, m, d, p, q son enteros positivos tales que $n = k + m$, p y q satisfacen que $q > p$ y B_k es una contante que se ha determinado de manera experimental. En [11] se ha propuesto establecer $B_k = 40$.

Es importante mencionar que en el esquema de firma propuesto en [11] y en el esquema estándar de firma $pqNTRUSign$ se realizan los algoritmos de firma y verificación, considerando dos distribuciones de polinomios, a saber distribución gaussiana discreta y distribución uniforme discreta para asegurar que se logre un muestreo de rechazo y aleatoriedad.

En esta tesis nos enfocaremos únicamente en la firma digital donde los polinomios siguen una distribución uniforme k-dimensional módulo q, U_q^k, que se estudia en [35].

La manera de establecer las claves de $NTRUSign$ es como sigue:

- Aldo selecciona de manera aleatoria polinomios $F(x) \in T(d+1, d), g(x) \in T(d+1, d)$, para generar los polinomios $F(x)$ y $g(x)$, Aldo puede usar el algoritmo 1 de la sección 6.5.2. Se establece $f(x) = pF(x)$.
- Enseguida calcula los polinomios $g^{-1}(x) \in R_p$ y $F_q(x) = f^{-1}(x) \in R_q$. En caso de que $g(x)$ y $f(x)$ no sean invertibles, se deben elegir nuevos polinomios $g(x)$ y $f(x)$ que sí tengan inverso multiplicativo en R_p y R_q, respectivamente.
- Luego, se verifican las condiciones $NORMF(f(x)) < B_k$ y $NORMF(g(x)) < B_k$. En caso de que alguna de las dos condiciones no se cumpla, Aldo deberá elegir nuevos polinomios $f(x)$ o $g(x)$ que sí cumplan lo anterior.
- Finalmente se calcula el producto $h(x) = g(x)F_q(x) \mod q$.

Según el esquema de firma propuesto por J. Hoffstein et al. en [11] se deben fijar dos constantes para determinar si una firma es aceptada o rechazada en el algoritmo de firma, las cuales son $B_t = 49$ y $B_m = 98$ cuando se trabaja con polinomios tal que sus coeficientes tienen una distribución uniforme.

La constante B_m es un parámetro que depende solo del parámetro d. Se determinan experimentalmente mediante un promedio de un gran número de muestras de $f(x)$ elegidas aleatoriamente con su n-ada de coeficientes en $\mathbb{Z}_p^n$ y del polinomio $a(x)$, que se define en la sección

6.5.6, elegido aleatoria y uniformemente, con coeficientes en $\mathbb{Z}_p^n$. Análogamente la constante B_t se determina como un promedio de un gran número de opciones de $g(x)$ y $a(x)$, como lo comenta J. Hoffstein et al., en [11].

Los parámetros y claves que ha generado Aldo para el esquema $NTRUSign$ quedan de la siguiente manera:

1. Los parámetros $n, k, m, p, q, B_t = 49$ y $B_m = 98$ son públicos.
2. Se deja como clave pública al polinomio $h(x)$.
3. La clave privada serán la pareja de polinomios $(f(x), g(x))$.

6.5.6. Algoritmo de firma digital con NTRU

El siguiente algoritmo calcula la firma digital $NTRUSign$ para un mensaje $\mu(x)$ utilizando una clave privada $(f(x), g(x))$.

Supongamos que Aldo desea firmar un mensaje o documento digital $\mu(x)$. Entonces, considera los polinomios seleccionados de la sección anterior, $h(x), f(x), g(x)$ y una sucesión aleatoria de bits, s, como fuente de aleatoriedad. Con ellos genera una firma $(m'(x), t'(x))$ para el mensaje $\mu(x)$.

Entrada: $(f(x), g(x), h(x), \mu(x), s)$, donde $(f(x), g(x))$ es una clave privada, $h(x)$ su correspondiente clave pública, el mensaje $\mu(x)$ está previamente convertido en forma binaria y s es una cadena aleatoria de bits.

1. Aldo genera el resumen del mensaje mediante una función hash aprobada por el Instituto Nacional de Estándares y Tecnología (NIST). Como ejemplo Aldo puede usar la función SHA-3-512 como función hash, para así obtener el resumen *msg_digest*.
2. **Repetir**:
   - 2.1 Aldo elige dos polinomios aleatorios uniformes. Por ejemplo, utilizando el algoritmo 2 de la sección 6.5.2, para obtener dichos polinomios con parámetros k y m, módulo $p = 2$, se toma como cadena aleatoria al resumen del mensaje *msg_digest* para obtener polinomios $m_p(x) \in R(p/2)$ y $t_p(x) \in R(p/2)$. Note que los polinomios $m_p(x)$ y $t_p(x)$ tienen exactamente un total de k coeficientes y t coeficientes respectivamente.
   - 2.2 Aldo elige un polinomio uniforme aleatorio de la misma manera como en el paso 2.1 anterior con parámetros n, módulo $A = \lfloor \frac{q}{2p} + \frac{1}{2} \rfloor$ y semilla la sucesión aleatoria s, para obtener un polinomio $r(x) \in R(A)$.
   - 2.3 Aldo calcula los siguientes polinomios:
     a) $m_0(x) = m_p(x) + pr(x)$.
     b) $t_0(x) = h(x)m_0(x) \mod q$, con $t_0(x) \in R(q/2)$.
     c) $a(x) = g^{-1}(x)(t_p(x) - t_0(x)) \mod p$, con $a(x) \in R(p/2)$.
     d) $m'(x) = m_0(x) + a(x)f(x)$.
     e) $t'(x) = t_0(x) + a(x)g(x)$.
3. **Hasta que**

$$||a(x)f(x)||_\infty \leq B_m,\ ||a(x)g(x)||_\infty \leq B_t,\ ||m'(x)||_\infty \leq \tfrac{q}{2} - B_m,\ ||t'(x)||_\infty \leq \tfrac{q}{2} - B_t.$$

4. **Salida:** $(m'(x), t'(x), \mu(x))$.

6.5.7. Algoritmo de verificación

La firma $(m'(x), t'(x))$ para el mensaje $\mu(x)$ obtenida por el algoritmo de firma anterior se verifica de la siguiente manera:

1. **Entrada:** $(m'(x), t'(x), \mu(x), h(x), s)$.
2. El usuario que desee verificar una firma primero debe elegir dos polinomios aleatorios uniformes mediante el algoritmo 2 de la sección 6.5.2, para obtener dichos polinomios con parámetros k y m, módulo $p = 2$, y cadena aleatoria al resumen del mensaje *msg_digest*, y así, obtener polinomios $m_p(x) \in R(p/2)$ y $t_p(x) \in R(p/2)$.

   Luego, verifica las siguientes condiciones:
3. Si $t'(x) \not\equiv h(x)m'(x)$ (mód q), envíe "no válida"si es así.
4. Si $||m'(x)||_\infty > q/2 - B_m$ o $||t'(x)||_\infty > q/2 - B_t$, envíe "no válida"si es así.
5. Si $(m'(x), t'(x)) \not\equiv (m'_p(x), t'_p(x))$ (mód p), envíe "no válida"si es así.
6. **Salida:** "Válida".

Un método de validación de pares de claves determina si un par de clave pública/clave privada que es candidato cumple con las restricciones para los pares de claves producidos por un método de generación de claves en particular.

Actualmente no existen métodos eficientes que proporcionen una validación de clave pública completa para el esquema anterior, aunque se esperaría que dicha validación garantice que se cumpla la relación $h(x) = g(x)F_q(x) \mod q$.

6.5.8. Parámetros

A continuación se presenta una tabla de valores específicos para los parámetros del esquema de firma $NTRUSign$, que brindan un nivel específico de seguridad de 269 bits de seguridad clásica y 149 bits de seguridad cuántica. Dichos parámetro difieren un poco del esquema estándar $pqNTRUSign$ debido a que el modelo de estimación es diferente, [11].

Parámetros	R	n	q	d	B_k, B_m, B_t	Tamaño de $h(x)$	Tamaño de $(m'(x), t'(x))$
$NTRUSign$	$\frac{\mathbb{Z}_q[\omega][x]}{(x^n+1)}$	1024	$2^{16}+1$	205	40, 98, 49	16384 bits	16384 bits

Tabla 6.1: Se muestra una elección adecuada de parámetros para el esquema de firma $NTRUSign$.

En la tabla anterior vemos que la longitud de la clave pública $h(x) = h_0 + h_1x + \cdots + h_{n-1}x^{n-1}$ debe ser de 16384 bits, ya que $h(x)$ está representado módulo q, así que cada coeficiente h_i está en el intervalo $[0, 2^{16}]$, por lo que, la longitud binaria de cada coeficiente h_i se puede expresar como cadenas de bits de longitud 16. Contando los n coeficientes de $h(x)$ tenemos que la longitud de $h(x)$ en bits es precisamente $n \cdot 16 = 1024 \cdot 16 = 16384$ bits.

Para la firma, vemos que los polinomios $m'(x)$ y $t'(x)$ tienen grado $k-1$ y $m-1$, respectivamente, donde $n = k + m$. La concatenación de los coeficientes de $m'(x)$ y $t'(x)$ es un

vector tamaño n, y con un razonamiento análogo al párrafo anterior es fácil ver que la firma $(m'(x), t'(x))$ es de longitud 16384 bits.

Considerando la observación 6.5.1 vemos que una práctica común para la clave pública $h(x)$ y la firma digital $(m'(x), t'(x))$ es que expresemos a dichos elementos en su forma binaria, es decir, en cadenas de bits. Para lograr dicha expresión podemos usar el algoritmo 4 (RE2BSP) de la sección 6.5.2 y expresar a $h(x)$ y a $(m'(x), t'(x))$ en cadenas de octetos. De manera viceversa, para pasar de cadenas de octetos a elementos de un anillo de polinomios podemos usar el algoritmo 9 (BS2REP) o los algoritmos 6 (RE2OSP) y 10 (OS2REP) de la misma sección ya mencionada.

6.6. Análisis de seguridad

Existen diversos ataques contra los esquemas de firma digital, por ejemplo, los de la sección 6.4, que son ataques con solo la clave pública, los ataques de transcripción etc. En el esquema de firma $NTRUSign$, que es un esquema basado en lattices, veremos que parte de la seguridad de este esquema sobre un lattice genérico se puede reducir al problema del vector más corto SVP sobre este lattice genérico.

Más específicamente, consideraremos dos tipos de ataques; el ataque de clave pública que intentan recuperar la clave secreta a partir de una clave pública, y el ataque de falsificación que intenta falsificar una firma sin tener acceso a la clave secreta. En esta sección daremos un método para proceder a realizar un ataque e intentar recuperar la clave secreta $f(x)$ y $g(x)$, a partir de la clave pública $h(x)$, que es similar al ataque para el criptosistema $NTRU$ de la sección 3.4, en el capítulo 3.

Los siguientes conceptos son algunas fortalezas en las que se basa la seguridad del esquema de firma $NTRUSign$, pero que a su vez son puntos clave para realizar ataques a este esquema si se consideran elecciones no adecuadas de parámetros, por ejemplo:

1. La seguridad de la clave pública se basa en el problema de solución de vectores cortos, con coordenadas en los enteros $\mathbb{Z}$.
2. La seguridad de falsificación se basa en el problema de aprendizaje con truncamiento.
3. La seguridad de transcripción es producida por muestreo de rechazo.

Mientras que, el esquema estándar $pqNTRUSign$ es una instancia eficiente de la firma con lattices modulares NTRU, con los siguientes supuestos de seguridad,

1. La seguridad de la clave pública se basa en la suposición 1 de $NTRU$, (ver 6.20).
2. La seguridad contra la falsificación se basa en el problema de aprendizaje con truncamiento sobre el lattice NTRU.
3. La seguridad de transcripción es producida por muestreo de rechazo.

En esta sección estudiaremos los tres puntos de seguridad mencionados anteriormente para el esquema de firma $NTRUSign$.

6.6.1. Ataque con solo la clave pública $h(x)$

Actualmente se tienen diferentes ataques contra lattices que combinan dos o más métodos de ataques comunes, a estos ataques se les conoce como ataques híbridos. Uno de los ataques

más conocidos contra lattices NTRU, es el ataque híbrido dado en [16], que combina ataques de reducción de lattices con ataques de encuentro en el medio.

Esencialmente la seguridad de la clave pública está determinada por la dificultad de resolver el problema de encontrar el único vector más corto en el lattice NTRU.

Por lo tanto, si un atacante potencial conoce la clave pública $h(x)$, puede emprender un ataque análogo al de la sección 3.4 del capítulo 3, y aplicar técnicas de reducción de lattices, como lo es el algoritmo LLL, para intentar recuperar vectores cortos que pueden ser la misma clave privada $(f(x), g(x))$ o rotaciones de ella misma.

Dado que en este ataque y en el ataque del lattices al criptosistema NTRU del capítulo 3, se considera el mismo lattice NTRU, con las mismas dimensiones, podemos deducir fácilmente que se tendrá como longitud objetivo τ, y una longitud esperada del vector distinto de cero más corto s, en L^{NT}, dadas por,

$$s = \sqrt{\frac{2n}{2\pi e}}(\lambda^n q^n)^{1/2n} = \sqrt{\frac{n}{\pi e}}\lambda^{1/2} q^{1/2} = \sqrt{\frac{\lambda n q}{\pi e}},$$

donde $2n$ es la dimensión del lattice L^{NT}, con determinante igual a $\lambda^n q^n$.

Al final de la sección 3.4.1 del capítulo 3 se brindan los argumentos necesarios para establecer que podemos tomar a la constante de equilibrio $\lambda = 1$. Por lo tanto, se tiene que,

$$s = \sqrt{\frac{nq}{\pi e}}.$$

Asumiendo que $\lambda = 1$, es fácil ver que la longitud objetivo será,

$$\tau = ||(f(x), g(x))||_2.$$

Por lo tanto, se dispone de los cálculos necesarios para emprender un ataque de reducción de lattices, con solo la clave pública $h(x)$ para el esquema de firma digital $NTRUSign$.

6.6.2. Ataque de falsificaciones

Notemos que los problemas de vectores más cortos y cercanos (SVP y CVP) se analizan utilizando la norma euclidiana y no la norma infinito, así que, escribiremos con la notación usual $||v||_2 = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$ para referirnos a la norma euclidiana del vector $v = (v_1, v_2, \ldots, v_n)$.

Uno de los mejores ataques de falsificaciones, además de derivar las claves secretas a partir de las claves públicas es el ataque de reducción de lattices que se muestra en [11], y que presentamos en esta sección.

La falsificación de una firma se puede lograr si se resuelve el problema del vector más corto aproximado asociado a la intersección de los lattices L^{NT} y $\mathbb{Z}^{2n}$. Por lo tanto, la tarea de falsificación de firmas se puede resolver encontrando un vector que cumpla con los requisitos de congruencia módulo p, y que esté lo suficientemente cerca del lattice de intersección ya mencionado para satisfacer los requisitos de norma del algoritmo de verificación en el esquema $NTRUSign$.

Proposición 6.6.1. *Sea $L_1 \subset \mathbb{Z}^r$ y $L_2 \subset \mathbb{Z}^r$ lattices de rango r, suponga que $t_1, t_2 \in \mathbb{Z}^r$ son vectores arbitrarios y sea $M = (L_1 + t_1) \cap (L_2 + t_2)$ la intersección de las traslaciones de L_1 y L_2. Hacemos las siguientes suposiciones:*

1. $\text{mcd}(\det(L_1), \det(L_2)) = 1$.

2. $t_1 \notin L_1$ *o* $t_2 \notin L_2$, *por lo que, en particular* $M \neq L_1 \cap L_2$.

Entonces se cumplen las siguientes condiciones:

a) $\det(L_1 \cap L_2) = \det(L_1)\det(L_2)$.

b) $M \neq \emptyset$.

c) Para cada $w_0 \in M$, *la función:*

$$L_1 \cap L_2 \to M, \qquad v \mapsto v + w_0,$$

es una biyección.

d) Sea $w_0 \in M$ *y suponga que* $w' \in M$ *es un vector distinto de cero más corto en* M. *Entonces* $w_0 - w'$ *resuelve el problema del vector más cercano en* $L_1 \cap L_2$, *para el vector* w_0 *(Esto es cierto para cualquier norma en* $\mathbb{Z}^r$, *por lo que, en particular es cierto para la norma euclidianas y la norma infinito).*

Demostración. La prueba del inciso **a)** es como sigue:

Sean $L_1 = L(B_1)$ y $L_2 = L(B_2)$, lattices de rango r, generadas por las bases B_1 y B_2, respectivamente. Luego, $L_1 + L_2 = L(B_1|B_2)$ es el lattices generado por la concatenación de las bases B_1 y B_2. Equivalentemente, podemos ver al lattice $L_1 + L_2$ como la suma de todos los pares de puntos de L_1 con L_2, es decir, $L_1 + L_2 = \{t_1 + t_2 | t_1 \in L_1, t_2 \in L_2\}$. Esta operación, $+$, puede verse como "dual"de la operación $\cap$, en lattices de rango completo, [9].

Como $L_1 \cap L_2 \subseteq L_1$, es decir, $L_1 \cap L_2$ es un lattices; entonces el cociente $L_2/(L_1 \cap L_2)$ está bien definido. Entonces por el teorema [[34], sección 3, teorema 3.6] se tiene que,

$$|L_2/(L_1 \cap L_2)| = det(L_1 \cap L_2)/det(L_2).$$

Por el segundo teorema del isomorfismos (considerando a los lattices como $\mathbb{Z}$-módulos), se tiene que,

$$L_2/(L_1 \cap L_2) \cong (L_1 + L_2)/L_1.$$

Entonces,

$$\det(L_1 \cap L_2)/\det(L_2) = \det(L_1)/det(L_1 + L_2)$$
$$\det(L_1 \cap L_2) = \frac{\det(L_1)\det(L_2)}{\det(L_1 + L_2)}.$$

Esto se cumple para cualquier lattice L_1 y L_2, tales que $L_1 \cap L_2$ y $L_1 + L_2$ son lattices.

Ahora, solo falta probar que $\det(L_1 + L_2) = 1$.

En textos sobre teoría de lattices, por ejemplo, en [9] y [24], se verifica que para lattices $L \subset \mathbb{Z}^r$, $\det(L)\mathbb{Z}^r \subseteq L \subseteq \mathbb{Z}^r$. También si $L \subset L'$, entonces $L + L_3 \subseteq L' + L_3'$. Aplicando los dos hechos anteriores a los lattices L_1 y L_2, tenemos que,

$$\det(L_1)\mathbb{Z}^r + \det(L_2)\mathbb{Z}^r \subseteq L_1 + L_2 \subseteq \mathbb{Z}^r.$$

Además, dado que $a\mathbb{Z}^r + b\mathbb{Z}^r = \text{mcd}(a,b)\mathbb{Z}^r$, y al usar nuestra hipótesis inicial de que $\text{mcd}(\det(L_1), \det(L_2)) = 1$, obtenemos que $\det(L_1)\mathbb{Z}^r + \det(L_2)\mathbb{Z}^r = \mathbb{Z}^r$, se sigue que,

$$\mathbb{Z}^r \subseteq L_1 + L_2 \subseteq \mathbb{Z}^r.$$

Por lo tanto, $\det(L_1 + L_2) = 1$.

b) La prueba del inciso b) es como sigue:

Hacemos $D_i = det(L_i)$, para $i = 1, 2$. Usaremos el hecho de que para cualquier lattice $L \subset \mathbb{Z}^r$, con determinante D, se cumple que $D\mathbb{Z}^r \subset L$. La hipótesis de que $\text{mcd}(D_1, D_2) = 1$, significa que podemos encontrar $x, y \in \mathbb{Z}$, tal que:

$$xD_1 + yD_2 = 1.$$

Establecemos,

$$e_1 = yD_2 = 1 - xD_1, \qquad e_2 = xD_1 = 1 - yD_2,$$

y consideramos el siguiente vector $t = e_1t_1 + e_2t_2$. Luego,

$$t - t_1 = (e_1 - 1)t_1 + e_2(t_2) = -xD_1t_1 + xD_1t_2 \in D_1\mathbb{Z}^r \subset L_1,$$

y similarmente,

$$t - t_2 = e_1t_1 + (e_2 - 1)t_2 = yD_2t_1 - yD_2t_2 \in D_2\mathbb{Z}^r \subset L_2.$$

Como el vector t está en M, entonces $M \neq \emptyset$.

c) La prueba del inciso c) es como sigue:

Para probar que la función $v \mapsto v + w_0$ es una biyección, demostraremos lo siguiente:

$$\text{Si } v \in L_1 \cap L_2, \text{ entonces } v + w_0 \in M.$$

$$\text{Si } w \in M, \text{ entonces } w - w_0 \in L_1 \cap L_2.$$

Para la condición $v \in L_1 \cap L_2 \Rightarrow v + w_0 \in M$, sabemos que $w_0 = v_1 + t_1 = v_2 + t_2$, con $v \in L_1$ y $v_2 \in L_2$, luego,

$$v + w_0 = \underbrace{(v + v_1)}_{\in L_1} + t_1 = \underbrace{(v + v_2)}_{\in L_2} + t_2,$$

así que, $v + w_0 \in M$.

Para la condición $w \in M \Rightarrow w - w_0 \in L_1 \cap L_2$, escribiremos al elemento dado, $w \in M$, como $w = v_1' + t_1 = v_2' + t_2$, con $v_1' \in L_1$ y $v_2' \in L_2$, por lo que,

$$w - w_0 = \underbrace{v_1' - v_1}_{\in L_1} = \underbrace{v_2' - v_2}_{\in L_2},$$

así que, $w - w_0 \in L_1 \cap L_2$.

d) La prueba del inciso d) es como sigue:

Sean w_0, w' elementos en M, tales que $||w'||_2 = \text{mín}_{w\in M\setminus\{0\}} ||w||_2$. Para facilitar la notación, escribiremos $v' = w_0 - w'$. Por el inciso c) de esta proposición, podemos suponer que $w' - w_0 \in L_1 \cap L_2$ y que $L_1 \cap L_2$ es un lattice, luego, estimamos lo siguiente,

$$\begin{aligned} ||v' - w_0||_2 &= ||w'||_2, \qquad \text{por definición de } v' \\ &= \min_{w\in M\setminus\{0\}} ||w||_2, \qquad \text{por definición de } w' \\ &= \min_{v\in(L_1\cap L_2)\setminus w_0} ||-v + w_0||_2, \quad \text{ya que el inciso } c) \text{ dice que } M = (L_1 \cap L_2) + w_0 \end{aligned}$$

Por lo tanto, si $w_0 \notin L_1 \cap L_2$, habremos probado que,

$$||v' - w_0||_2 = \min_{v\in L_1\cap L_2} ||v - w_0||_2,$$

que es el resultado deseado.

Supongamos que $w_0 \in L_1 \cap L_2$ y ya que también $w_0 \in M = (L_1 + t_1) + (L_2 + t_2)$ por hipótesis, podemos escribir,

$$w_0 = v_1 + t_1 \quad \text{y} \quad w_0 = v_2 + t_2,$$

con $v_1 \in L_1$ y $v_2 \in L_2$.

Pero entonces $t_1 = w_0 - v_1 \in L_1$ y $t_2 = w_0 - v_2 \in L_2$, lo cual contradice la suposición inicial 2) sobre t_1 y t_2. Por lo tanto, $w_0 \notin L_1 \cap L_2$. □

Recordemos dos cantidades que están asociadas a los problemas de lattices.

Heurística 1. La heurística gaussiana dice que el tamaño respecto a la norma euclidiana de una solución del problema SVP o CVP en un lattice aleatorio L, de dimensión razonablemente grande es aproximadamente,

$$\gamma(L) = \sqrt{\frac{dim(L)}{2\pi e}} det(L)^{1/dim(L)}. \tag{6.2}$$

En otras palabras para la "mayoría"de lattices L y la "mayoría"de los vectores objetivo v_0 se tiene que,

$$\text{mín}_{v\in L\setminus\{0\}} ||v||_2 \approx \gamma(L) \quad \text{y} \quad \text{mín}_{v\in L} ||v - v_0||_2 \approx \gamma(L).$$

Heurística 2. Sea $L \subset \mathbb{Z}^n$ un lattice para el cual queremos resolver el problema τ-aprox-SVP o τ-aprox-CVP. En otras palabras, sea $v_0 \in \mathbb{Z}^n$, y suponga que queremos encontrar un vector $v \in L$ que satisfaga,

$$0 < ||v||_2 \leq \tau \quad o \quad ||v - v_0||_2 \leq \tau.$$

Llamamos τ a *la longitud objetivo del problema*. El defecto gaussiano del problema es la relación,

$$\rho = \rho(L, \tau) = \frac{\tau}{\gamma(L)}. \tag{6.3}$$

Sea $0 < \delta < 2$. La heurística δ-LLL, que ha sido confirmada en numerosos experimentos, dice que resolver el problema τ-aprox-SVP o τ-aprox-CVP es exponencialmente difícil en función de $dim(L)$, siempre que el defecto gaussiano $\rho(L, \tau)$ no sea más que un múltiplo pequeño de $dim(L)^\delta$. Un desarrollo completo sobre la heurística se puede encontrar en [10] y en [2].

Consideremos el problema de falsificar una firma, el falsificador necesita encontrar un vector $(m,t) \in L^{ET}$, que satisfaga lo siguiente,

$$(m(x), t(x)) = (m_p(x), t_p(x)) \quad \text{mód } p, \qquad \ldots\ldots \text{ condición de congruencia.} \tag{6.4}$$

$$||m(x)||_\infty \leq \frac{q}{2} - B_m, \qquad \ldots\ldots \text{ condición de norma.} \tag{6.5}$$

$$||t(x)||_\infty \leq \frac{q}{2} - B_t, \qquad \ldots\ldots \text{ condición de norma.} \tag{6.6}$$

Los vectores m_p, t_p están en $R(p/2)$, por lo que, la condición de congruencia anterior puede formularse diciendo que la condición objetivo (m,t) está en la traslación del lattice $p\mathbb{Z}^{2n}$ por el vector (m_p, t_p). Por lo tanto, el falsificador está buscando un vector corto en relación con la norma infinito en la intersección siguiente,

$$(m,t) \in L^{NT} \cap (p\mathbb{Z}^{2n} + (m_p, t_p)).$$

En la sección 3.4 del capítulo 3, vimos que $\det(L^{NT}) = q^n$, siempre que se tenga como constante de equilibrio $\lambda = 1$. Además, es fácil ver que $\det(p\mathbb{Z}^{2n}) = p^{2n}$. Las cantidades q^n y p^{2n} son primos relativos, pues p y q se eligen tales que sean primos relativos. Entonces podemos usar la proposición 6.6.1, inciso a), para concluir que,

$$\det(L^{NT} \cap p\mathbb{Z}^{2n}) = q^n p^{2n}.$$

Luego, la proposición 6.6.1 inciso d) nos dice que encontrar un vector corto en la intersección $L^{NT} \cap (p\mathbb{Z}^{2n} + (m_p, t_p))$ es equivalente a resolver el problema τ-aprox-CVP en el lattice de intersección $L^{NT} \cap p\mathbb{Z}^{2n}$. Un cálculo rutinario de la heurística gaussiana para $L^{NT} \cap p\mathbb{Z}^{2n}$ nos permite ver que,

$$\gamma(L^{NT} \cap p\mathbb{Z}^{2n}) = \sqrt{\frac{n}{\pi e}}(q^n p^{2n})^{1/2n} = \sqrt{\frac{p^2 qn}{\pi e}}.$$

Solo queda estimar la longitud objetivo. El criterio de rechazo en el algoritmo de firma $NTRUSign$ dice que una firma válida $(m'(x), t'(x))$ tiene una norma superior de máximo $\frac{q}{2} - \min(B_m, B_t)$. Por lo tanto, en particular, una firma válida satisface el límite de la norma euclidiana, que se describe a continuación:

Si $(m'(x), t'(x))$ es una firma válida, entonces debe satisfacer que:

$$||m'(x)||_\infty = \max_i |m_i| \leq q/2 - B_m \quad \text{y} \quad ||t'(x)||_\infty = \max_i |t_i| \leq q/2 - B_t.$$

Entonces, tenemos que,

$$\begin{aligned} ||(m'(x), t'(x))||_2 &= \sqrt{m_0^2 + m_1^2 + \cdots + m_{n-1}^2 + t_0^2 + t_1^2 + \cdots + t_{n-1}^2} \\ &\leq \sqrt{n\left(\frac{q}{2} - B_m\right)^2 + n\left(\frac{q}{2} - B_t\right)^2} \\ &\leq \sqrt{\left(\frac{q}{2} - \min\{B_m, B_t\}\right)^2}\sqrt{2n} \\ &= \left(\frac{q}{2} - \min\{B_m, B_t\}\right)\sqrt{2n}. \end{aligned}$$

Por lo tanto, el límite de la norma euclidiana es,

$$||(m'(x), t'(x))||_2 \leq \left(\frac{q}{2} - \min\{B_m, B_t\}\right)\sqrt{2n}. \tag{6.7}$$

Pero no todo vector en el lattice L^{NT} que satisfaga la condición de norma euclidiana en la desigualdad (6.7) y la condición de congruencia (6.4), será una firma válida. Para simplificar o hacer que sea más fácil para un falsificador potencial realizar una falsificación, supongamos que solo se necesita la condición de norma (6.7), en lugar de la condición de norma (6.5 y 6.6).

Además, poniéndonos a favor del falsificador, supondremos que $B_m = B_t = 0$, de modo que solo se necesita encontrar un vector en $R(\frac{q}{2}) \times R(\frac{q}{2})$. Esto nos da una longitud objetivo $\tau = ||(m'(x), t'(x))||_2 = q\sqrt{n/2}$.

Entonces, el defecto gaussiano para nuestro problema aproximado CVP es,

$$\rho = \frac{q\sqrt{n/2}}{\sqrt{p^2qn/\pi e}} = q\sqrt{\frac{\pi e}{2p^2q}} \approx n64\sqrt{\frac{\pi e}{2p^2q}}, \tag{6.8}$$

que es aproximadamente un múltiplo pequeño de $dim(L^{NT} \cap\ p\mathbb{Z}^{2n}) = 2n$, así que la heurística δ-LLL dice que resolver el problema aproximado CVP asociado es un problema (computacionalmente) difícil.

Este análisis funciona tanto para polinomios que siguen una distribución gaussiana como para los que siguen una distribución uniforme. J. Hoffstein en [7] proporciona un mejor ajuste de parámetros para el esquema $NTRUSign$, que recordemos está basado solo en polinomios uniformes, de modo que si consideramos únicamente distribución uniforme, tendremos los siguientes parámetros.

El lattice de intersección $L^{NT} \cap\ p\mathbb{Z}^{2n}$, es generado por las filas de la matriz,

$$B^{NT} = \begin{bmatrix} 0 & pqI_n \\ pI_n & pH \end{bmatrix},$$

para alguna matriz H apropiada, generada por una clave pública $h(x)$. También suponemos que este lattice se comporta como un lattice aleatorio.

En el esquema de firma $NTRUSign$, cada coordenada de los vectores m' y t' se distribuyen aproximadamente de manera aleatoria y uniforme entre los intervalos:

$$\left(-\frac{q}{2} + B_m, \frac{q}{2} - B_m\right] \quad \text{y} \quad \left(-\frac{q}{2} + B_t, \frac{q}{2} - B_t\right],$$

respectivamente.

Ignorando las constantes B_i, es decir, considerando que $B_m = B_t = 0$, el coeficiente cuadrático medio de los coeficientes de (m', t') será aproximadamente:

$$\frac{1}{q}\int_{-q/2}^{q/2} x^2 dx = \frac{q^2}{12}.$$

Así, los vectores m' y t' tendrán norma euclidiana $||m'||_2^2 \approx ||t'||_2^2 \approx q^2n/12$. Por lo tanto,

$$\rho = \frac{\text{longitud objetivo}}{\text{longitud heuristica Gaussiana}} = \frac{\sqrt{q^2n/6}}{\sqrt{np^2q/\pi e}} = q\sqrt{\frac{\pi e}{6p^2q}} = \sqrt{\frac{q\pi e}{6p^2}}. \tag{6.9}$$

Note que se han propuesto dos posibles valores del defecto gaussiano, (6.8) y (6.9). En la práctica es recomendable usar (6.9), ya que la selección de parámetros de la sección 6.5.5 están mejor ajustados para polinomios uniformes.

Por lo tanto, algún usuario potencial puede emplear el ataque anterior y usar técnicas de reducción de lattices para intentar obtener firmas falsas para ciertos mensajes $\mu_i(x)$, pero dicho ataque puede resultar poco exitoso cuando tenemos parámetros adecuados, como lo especifica la tabla 6.1. La elección de buenos parámetros mantiene una buena seguridad cuando se implementa este esquema de firma digital para firmar documentos electrónicos.

6.6.3. Seguridad de transcripciones

En esta sección se prueba que bajo una suposición razonable, una transcripción de firmas generada con el algoritmo de firma digital $NTRUSign$ no contiene información que no esté ya disponible para alguien que conoce la clave pública $h(x)$, es decir, que no se tiene información adicional sobre las firmas creadas, más que la información que se puede generar a partir de la clave pública $h(x)$. Esto se logrará probando que si un usuario obtiene firmas válidas con el esquema $NTRUSign$, entonces estas firmas se distribuyen de manera aleatoria y uniforme en el lattice de norma acotada $L^{NT}\left(\frac{q}{2} - B_m, \frac{q}{2} - B_t\right)$.

Considerando los tipos de ataques a un esquema de firma digital de la sección 6.4, entenderemos por un ataque de transcripción a algún método para recuperar la clave privada a partir de una larga lista (transcripción) de firmas de la forma,

$$(firma\ válida_i,\ documento\ hash\ i), \quad i = 1, 2, \ldots \tag{6.10}$$

Veremos un resultado, que se encuentra en [11], el cual afirma que la distribución de firmas es uniforme sobre el lattice $(m_p, t_p) + p\mathbb{Z}^{2n}$. Además, se prueba que un usuario que conoce la clave pública $h(x)$ solo puede producir una transcripción de pares como en (6.10), que es estadísticamente indistinguible de una transcripción análoga producida por el mismo esquema firma $NTRUSign$ y una clave secreta $(f(x), g(x))$.

Iniciamos con el análisis de una transcripción creada, usando el esquema de firma $NTRUSign$ con clave secreta $(f(x), g(x))$. La condición de muestreo de rechazo que proporciona este esquema permite probar que las firmas resultantes tengan una distribución uniforme en una cierta región de firmas permitidas.

Supongamos que a partir de la función hash que se utiliza en el esquema $NTRUSign$ se generan varios documentos hash de la forma,

$$(m_p(x), t_p(x)) \in R(p/2) \times R(p/2),$$

los cuales se distribuyen uniformemente en $R(p/2) \times R(p/2)$. Esta hipótesis es posible ya que la generación de los polinomios se hace de manera aleatoria con el algoritmo 2 del a sección 6.5.2.

Consideramos los pasos 2.2 y 2.3 del algoritmo de firma $NTRUSign$ para definir una función de firma $(m'(x), t'(x)) = \sigma'(f(x), g(x), m_p(x), t_p(x), r(x))$. Por lo tanto, σ' es una función,

$$\sigma' : \underbrace{pR(3/2) \times R(p/2)}_{clave\ privada} \times \underbrace{R(p/2) \times R(p/2)}_{documento\ hash} \times \underbrace{R(\lfloor q/2p - 1/2 \rfloor)}_{elemento\ aleatorio\ r(x)} \to \underbrace{L^{NT}(q/2 + B_m, q/2 + B_t)}_{firma\ potencial},$$

dada explícitamente por:

$$\sigma'(f(x), g(x), m_p(x), t_p(x), r(x)) = (m_0(x) + a(x)f(x), t_0(x) + a(x)g(x)), \tag{6.11}$$

donde,

$$m_0(x) = m_p(x) + pr(x). \tag{6.12}$$

$$t_0(x) = h(x)m_0(x), \qquad \text{con } t_0(x) \in R(q/2). \tag{6.13}$$

$$a(x) = g^{-1}(x)(t_p(x) - t_0(x)), \qquad \text{con } a(x) \in R(p/2). \tag{6.14}$$

Escribiremos,

$$\Omega' = pR(3/2) \times R(p/2) \times R(p/2) \times R(p/2) \times R(\lfloor q/2p - 1/2 \rfloor),$$

para representar el dominio de la función σ'.

Ahora, introduciremos el muestreo de rechazo. Sea $\Omega = \{f(x), g(x), m_p(x), t_p(x), r(x)\}$ un subconjunto de Ω', tal que:

1. $(m'(x), t'(x)) := \sigma'(f(x), g(x), m_p(x), t_p(x), r(x)) = (m_0(x) + a(x)f(x), t_0(x) + a(x)g(x))$.
2. $||m'(x)||_\infty \leq q/2 - B_m, \qquad ||t'(x)||_\infty \leq q/2 - B_t$.
3. $||a(x)f(x)||_\infty \leq B_m, \qquad ||a(x)g(x)||_\infty \leq B_t$.

La restricción de σ' a Ω, la denotaremos por σ. Entonces, σ es una función,

$$\sigma : \Omega \to L^{NT}(q/2 - B_m,\ q/2 - B_t).$$

Establecemos la siguiente notación, $A = \lfloor \frac{q}{2p} + \frac{1}{2} \rfloor$. Por lo tanto, la elección del polinomio $r(x)$ en el paso 2.2 del algoritmo de firma $NTRUSign$ se toma de manera aleatoria y uniforme en el conjunto $R(A)$. A continuación estudiaremos dos resultados importantes que son útiles para demostrar la proposición 6.6.3, la cual dice que toda firma válida para el documento hash $(m_p(x), t_p(x))$ tiene el mismo número de preimagenes en el conjunto $R(A)$.

Proposición 6.6.2. *El esquema de firma NTRUSign produce firmas que son verificadas como válidas por el respectivo algoritmo de verificación.*

Demostración. Sea $(m'(x), t'(x), \mu(x))$ una firma generada por el algoritmo de firma $NTRUSign$ para el mensaje $\mu(x)$.

Verificaremos que $t'(x) \equiv h(x)m'(x) \pmod q$. Por definición de $t'(x)$ y $t_0(x)$, tenemos que,

$$t'(x) = t_0(x) + a(x)g(x) = h(x)m_0(x) + a(x)g(x).$$

Como $h(x) = F_q(x)g(x) \mod q$, entonces tenemos que $h(x)f(x) = g(x) \mod q$, y al sustituir a $g(x)$ en las expresiones de $t'(x)$ tenemos que,

$$\begin{aligned} t'(x) &= t_0(x) + a(x)g(x) = h(x)m_0(x) + a(x)g(x) \\ &\equiv h(x)m_0(x) + h(x)a(x)f(x) \ (\text{m}\acute{o}d\ q) \\ &\equiv h(x)[m_0(x) + a(x)f(x)] \ (\text{m}\acute{o}d\ q) \\ &\equiv h(x)m'(x) \ (\text{m}\acute{o}d\ q). \end{aligned}$$

Ahora verificaremos que $||m'(x)||_\infty \leq q/2 - B_m$ y $||t'(x)||_\infty \leq q/2 - B_t$. El paso 3 del algoritmo de firma del esquema $NTRUSign$ asegura que las normas anteriores siempre se satisfacen, pues es uno de los requisitos para que el algoritmo de firma termine. Por lo tanto, se satisfacen las dos desigualdades anteriores.

Finalmente probaremos que $(m'(x), t'(x)) \equiv (m_p(x), t_p(x))$ (m*ód* p).

Para $m'(x) \equiv m_p(x)$ (m*ód* p), tenemos,

$$\begin{aligned} m'(x) &= m_0(x) + a(x)f(x), \qquad \text{por definición de } m'(x) \\ &\equiv m_p(x) + pr(x) + a(x)f(x) \text{ (m\'od } p), \qquad \text{al sustituir } m_0(x) \\ &\equiv m_p(x) \text{ (m\'od } p), \qquad \text{al aplicar la reducción} \quad \text{mód } p. \end{aligned}$$

Para $t'(x) \equiv t_p(x)$ (m*ód* p), tenemos,

$$\begin{aligned} t'(x) &= t_0(x) + a(x)g(x), \qquad \text{por definición de } t'(x) \\ &\equiv t_0(x) + g^{-1}(x)(t_p(x) - t_0(x))g(x) \text{ (m\'od } p), \qquad \text{al sustituir } a(x) \\ &\equiv t_0(x) + t_p(x) - t_0(x) \text{ (m\'od } p), \qquad \text{al operar } g^{-1}(x) \text{ con } g(x) \\ &\equiv t_p(x) \text{ (m\'od } p). \end{aligned}$$

Se sigue que las firmas producidas por el esquema $NTRUSign$ se verifican como válidas por el algoritmo de verificación correspondiente. □

Denotemos por $\Sigma(f(x), g(x), m'(x), t'(x))$ a la colección,

$$\Sigma(f(x), g(x), m'(x), t'(x)) = \left\{ r(x) \in R(A) : \sigma(f(x), g(x), m_p(x), t_p(x), r(x)) = (m'(x), t'(x)) \right\}. \tag{6.15}$$

La clave para contar el tamaño del conjunto $\Sigma(f(x), g(x), m'(x), t'(x))$ es la biyección descrita en el siguiente lema.

Lema 6.6.1. *Sea* $C = \{b(x) \in R(p/2) : ||b(x)f(x)||_\infty \leq B_m \text{ y } ||b(x)g(x)||_\infty \leq B_t\}$ *y sea* $(m'(x), t'(x)) \in L^{NT}(q/2 - B_m, q/2 - B_t)$*, que satisface* $(m'(x), t'(x)) \equiv (m_p(x), t_p(x))$ (m*ód* p)*.*

Entonces la siguiente función es una biyección de conjuntos:

$$\begin{aligned} \Phi : C &\to \Sigma(f(x), g(x), m'(x), t'(x)), \\ b &\mapsto \frac{m'(x) - m_p(x)}{p} - b(x)\frac{f(x)}{p}. \end{aligned} \tag{6.16}$$

Demostración. Notemos que:

$$\begin{aligned} m'(x) - m_p(x) &= m_0(x) + a(x)f(x) - m_p(x), \qquad \text{por definición } m'(x) \\ &= m_p(x) + pr(x) + a(x)f(x) - m_p(x), \qquad \text{por definición } m_0(x) \\ &= pr(x) + a(x)f(x) \\ &= pr(x) + a(x)pF(x), \qquad \text{por definición } f(x) \\ &= p[r(x) + a(x)F(x)]. \end{aligned}$$

Entonces, los coeficientes de $m'(x) - m_p(x)$ son múltiplos de p, y de manera similar $f(x) \in pR(3/2)$ tiene coeficientes divisibles por p. Además, vemos que el polinomio del lado derecho en (6.16) tiene todos sus coeficientes en $\mathbb{Z}$.

Ahora, necesitamos verificar que $\Phi(b(x)) \in \Sigma(f(x), g(x), m'(x), t'(x))$, es decir, debemos probar que $\Phi(b(x)) \in R(A)$ y $\sigma(f(x), g(x), m_p(x), t_p(x), \Phi(b(x))) = (m'(x), t'(x))$. Tenemos que:

$$\begin{aligned}
||\Phi(b(x))||_\infty &= \left\| \frac{m'(x) - m_p(x) - b(x)f(x)}{p} \right\|_\infty \\
&\leq \frac{1}{p}\left[\left\|m'(x)\right\|_\infty + \left\| - m_p(x) - b(x)f(x)\right\|_\infty \right], \qquad \text{por la desigualdad del triángulo} \\
&\leq \frac{1}{p}\left[\left\|m'(x)\right\|_\infty + \left\|m_p(x)\right\| + \left\|b(x)f(x)\right\|_\infty \right], \qquad \text{por la desigualdad del triángulo} \\
&\leq \frac{1}{p}\left[\frac{q}{2} - B_m + \frac{p}{2} + B_m \right], \qquad \text{por que } ||m'(x)||_\infty \leq \frac{q}{2} - B_m,\ m_p(x) \in R(p/2), \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad y\ b(x) \in C \\
&= \frac{q}{2p} + \frac{1}{2}.
\end{aligned}$$

Entonces, $||\Phi(b(x))||_\infty \leq \frac{q}{2p} + \frac{1}{2}$ y aplicando la función piso de ambos lados de esta desigualdad, tenemos que,

$$||\Phi(b(x))||_\infty \leq \lfloor \frac{q}{2p} + \frac{1}{2} \rfloor = A.$$

Por lo tanto, se cumple que $\Phi(b(x)) \in R(A)$.

Luego, usamos las cuatro fórmulas (6.11) - (6.14) para calcular la firma,

$$\begin{aligned}
&\sigma(f(x), g(x), m_p(x), t_p(x), \Phi(b(x))) : \\
m_0(x) = m_p(x) + p\Phi(b(x)) &= m_p(x) + p\left(\frac{m'(x) - m_p(x)}{p} - b(x)\frac{f(x)}{p} \right) \\
&= m'(x) - b(x)f(x). \qquad (6.17)
\end{aligned}$$

$$\begin{aligned}
t_0(x) &= h(x)m_0(x) \pmod q \\
&\equiv h(x)(m'(x) - b(x)f(x)) \ (\text{m}\acute{o}d\ q) \\
&\equiv h(x)m'(x) - b(x)g(x) \ (\text{m}\acute{o}d\ q), \qquad \text{ya que } h(x) \equiv F_q(x)g(x) \pmod q \\
&\equiv t(x) - b(x)g(x) \ (\text{m}\acute{o}d\ q), \qquad \text{ya que } \ (m'(x), t'(x)) \in L^{NT}. \qquad (6.18)
\end{aligned}$$

Como $(m'(x), t'(x)) \in L^{NT}(q/2 - B_m, q/2 - B_t),\ b(x) \in C$ y por la desigualdad del triángulo tenemos que,

$$||m_0(x)||_\infty \leq ||m'(x)||_\infty + ||b(x)f(x)||_\infty = q/2 - B_m + B_m = q/2.$$

$$||t_0(x)||_\infty \leq ||t'(x)||_\infty + ||b(x)g(x)||_\infty = q/2 - B_t + B_t = q/2.$$

Por lo tanto, podemos concluir que la relación (6.18) es una igualdad y no solo una congruencia.

Continuando con el cálculo de $\sigma(f(x), g(x), m_p(x), t_p(x), \Phi(b(x)))$, usamos las ecuaciones (6.19) y (6.18) para calcular lo siguiente,

$$\begin{aligned}
a(x) &= g^{-1}(x)(t_p(x) - t_0(x)) \ (\text{mód } p) \\
&\equiv g^{-1}(x)t_p(x) - g^{-1}(x)t_0(x) \ (\text{mód } p) \\
&\equiv g^{-1}(x)t_p(x) - [g^{-1}t(x) - b(x)g^{-1}(x)g(x)] \ (\text{mód } p) \\
&\equiv b(x) \ (\text{mód } p).
\end{aligned}$$

Note que $t'(x) \equiv t_p(x)$ (mód p), por (6.19). Dado que tanto $a(x)$ como $b(x)$ están en $R(p/2)$, esto nos dice que $a(x) = b(x)$.

Ahora usamos la ecuación (6.11) y calculamos,

$$\begin{aligned}\sigma(f(x), g(x), m_p(x), t_p(x), \Phi(b(x))) &= (m_0(x) + a(x)f(x), t_0(x) + a(x)g(x)), \quad \text{por def. de } \sigma \\ &= (m'(x) - b(x)f(x) + a(x)f(x), t'(x) - b(x)g(x) + a(x)g(x)) \\ &= (m'(x), t'(x)), \qquad \text{pues } a(x) = b(x).\end{aligned}$$

En la penúltima igualdad anterior se utilizó (6.17) y (6.18). Por lo tanto, directamente de la definición (6.15) del conjunto $\Sigma(f(x), g(x), m'(x), t'(x))$, vemos que,

$$\Phi(b(x)) \in \Sigma(f(x), g(x), m'(x), t'(x)).$$

Por otro lado, tomamos un polinomio $r(x) \in \Sigma(f(x), g(x), m'(x), t'(x))$ y calculamos cuántos polinomios $b(x)$ están en el conjunto C, tales que se satisface $\Phi(b(x)) = r(x)$.

Dado que todos los coeficientes de los polinomios $m'(x) - m_p(x)$ y $f(x)$ son divisibles por p, para facilitar la notación escribiremos,

$$m'(x) - m_p(x) = pS(x) \qquad y \qquad f(x) = pF(x),$$

con $S(x)$ y $F(x)$ polinomios adecuados.

Recordemos que por hipótesis el polinomio $F(x)$ es invertible módulo p. Luego, tenemos que,

$$\begin{aligned}\Phi(b(x)) = r(x) &\iff S(x) - b(x)F(x) = r(x) \\ &\iff b(x) \equiv F^{-1}(x)(S(x) - r(x)) \text{ (mód } p),\end{aligned}$$

con $||b(x)||_\infty \leq \frac{p}{2}$.

Notemos que los polinomios $F^{-1}(x), S(x), r(x)$ son únicos y así el polinomio dado por $F^{-1}(x)(S(x) - r(x))$ módulo p es único. Por lo tanto, hay exactamente un valor de $b(x) \in C$ que satisface $\Phi(b(x)) = r(x)$, es decir, que el polinomio $b(x)$ es el único elemento del conjunto C que es congruente con $F^{-1}(x)(S(x) - r(x))$ módulo p. Esto prueba que la función Φ es biyectiva, lo que concluye la prueba de este lema. □

Proposición 6.6.3. *La función de firma σ tiene la siguiente propiedad: Para elementos fijos:*

$$\textit{clave privada} \quad (f(x), g(x)) \in pR \times R, \ \textit{y}$$

$$\textit{documento hash} \quad (m_p(x), t_p(x)) \in R(p/2) \times R(p/2),$$

la salida de σ, cuando se consulta sobre el polinomio uniforme aleatorio $r(x) \in R(A)$ se distribuye uniformemente sobre el conjunto,

$$\left\{(m'(x), t'(x) \in L^{NT}(q/2 - B_m, q/2 - B_t) : (m'(x), t'(x)) \equiv (m_p(x), t_p(x)) \text{ (mód } p)\right\},$$

de firmas válidas para el hash $(m_p(x), t_p(x))$.

De manera equivalente el tamaño del conjunto,

$$\{r(x) \in R(A) : \sigma(f(x), g(x), m_p(x), t_p(x), r(x)) = (m_p(x), t_p(x))\},$$

es igual para todos los $(m(x), t(x)) \in L^{NT}(q/2 - B_m, q/2 - B_t)$ que satisfacen

$$(m'(x), t'(x) \equiv (m_p(x), t_p(x)) \text{ (mód } p)).$$

Demostración. Por la proposición 6.6.2 sabemos que $\sigma(f(x), g(x), m_p(x), t_p(x), r(x))$ es congruente con $(m_p(x), t_p(x))$ m*ód* p, para cualquier evaluación de σ. Por lo tanto, se tiene que la probabilidad generar la firma $(m'(x), t'(x))$ si $(m'(x), t'(x)) \not\equiv (m_p(x), t_p(x))$ (m*ód* p) es cero. Así que, asumiremos de ahora en adelante que,

$$(m'(x), t'(x)) \equiv (m_p(x), t_p(x)) \text{ (m\'od } p), \tag{6.19}$$

para cualquier evaluación de σ.

El polinomio $r(x) = r_0 + r_1 x + \cdots + r_{n-1}x^{n-1}$ utilizado para generar una firma se elige de manera aleatoria y uniforme sobre el conjunto $R(A)$. Entonces, cualquier coeficiente r_i de $r(x)$ cumple que $|r_i| \leq A$, lo cual quiere decir que hay exactamente A distintas posibilidades de que r_i tome un valor entero positivo, $-A$ distintas posibilidades de que tome un valor entero negativo, y también la posibilidad de que $r_i = 0$, por lo que, cada coeficiente de $r(x)$ puede ser seleccionado de $(2A + 1)$ maneras distintas. Luego, aplicando el principio de la multiplicación tendremos $(2A + 1)^n$ opciones posibles para seleccionar el polinomio $r(x)$.

Por lo tanto, la probabilidad de obtener $(m'(x), t'(x))$ como firma en $(m_p(x), t_p(x))$ es igual al número de elementos del conjunto $\Sigma(f(x), g(x), m'(x), t'(x))$ entre $(2A + 1)^n$, donde $\Sigma(f(x), g(x), m'(x), t'(x))$ denota a la colección de conjuntos mencionada en la ecuación (6.15).

Entonces, para todas la firmas tales que $(m'(x), t'(x)) \equiv (m_p(x), t_p(x))$ (mód p), calculamos la siguiente probabilidad:

$$\begin{aligned} &P_{r(x)\longleftarrow R(A)}\left(\text{firma es } (m'(x), t'(x)) \mid \text{se tiene clave } (f(x), g(x)) \text{ y hash } (m_p(x), t_p(x))\right) \\ &\quad = \frac{\#\Sigma(f(x), g(x), m'(x), t'(x))}{\#R(A)} \\ &\quad = \frac{\#C}{\#R(A)}, \end{aligned}$$

donde la penúltima igualdad se sigue del lema 6.6.1. Esto completa la demostración de esta proposición. □

Para dar una prueba más completa de la seguridad de transcripciones, necesitamos una versión un poco más fuerte de la proposición 6.6.3, por lo que, se establece la siguiente proposición.

Proposición 6.6.4. *La distribución de firmas, con NTRUSign, al consultar σ sobre el hash del mensaje $(m_p(x), t_p(x)) \in R(p/2) \times R(p/2)$ uniformemente aleatorio, es indistinguible de la distribución uniforme sobre el lattice $L^{NT}(q/2 - B_m, q/2 - B_t)$.*

La proposición anterior es una consecuencia inmediata de la proposición 6.6.3 bajo el supuesto de que para cualquier $h(x)$ dada, el número de vectores del lattice de norma acotada, $L^{NT}(q/2 - B_m, q/2 - B_t)$, en cada clase lateral $p\mathbb{Z}^{2n}$ es esencialmente constante, es decir, que los elementos del lattice $L^{NT}(q/2 - B_m, q/2 - B_t) \cap (m_p, t_p) + p\mathbb{Z}^{2n}$ se mantienen distribuidos uniformemente.

La proposición anterior no se cumple para todos los lattices. Por ejemplo, con $h(x) = 1$, podemos obtener la siguiente matriz, que genera al lattice L^{NT},

$$B^{NT} = \left[\begin{array}{cccc|cccc} 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 1 \\ \hline 0 & 0 & \cdots & 0 & q & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & q & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & q \end{array}\right]_{2n\times 2n}$$

De está matriz podemos deducir que el lattice L^{NT} tiene rango n, además vemos que por definición $h(x) = F_q(x)g(x) \mod q$ y así,

$$h(x) = 1 = F_q(x)g(x) \quad \text{mód } q.$$

Además, como $F_q(x) = f^{-1}(x)$, tenemos que $f(x) = g(x) \in pR(3/2)$. Entonces, para encontrar la clave privada $(f(x), g(x))$, basta con encontrar a $f(x) \in R(3/2)$. Esto se reduce a encontrar un vector en $(m_p, t_p) + p\mathbb{Z}^{2n}$, el cual cuenta con solo p^n clases laterales distintas, por lo que, el esquema $NTRUSign$ puede quedar vulnerable.

Sin embargo J. Hoffstein en [11] propone que es probable que bajo el esquema de firma $NTRUSign$ se cumpla lo siguiente:

Suposición 1. Existen constantes reales c, ϵ tal que $\epsilon = 1/n^{O(1)}$ y para todo hash $(m_p(x), t_p(x)) \in R(p/2) \times R(p/2)$,

$$(1 - \epsilon)c \leq |L^{NT}(q/2 - B_m, q/2 - B_t) \cap [(m_p, t_p) + p\mathbb{Z}^{2n}]| \leq (1 + \epsilon)c. \quad (6.20)$$

Observación 6.6.1. *La suposición 1 nos dice que los elementos en el lattice de intersección $L^{NT}(q/2-B_m, q/2-B_t)\cap[(m_p, t_p)+p\mathbb{Z}^{2n}]$ mantienen una distribución uniforme, para cualquier documento hash $(m_p(x), t_p(x))$, lo cual hace resistente al esquema $NTRUSign$ ante los ataques de transcripciones.*

Concluimos esta sección señalando que cualquier usuario con acceso a $h(x)$ puede muestrear la distribución uniforme en $L^{NT}(q/2 - B_m, q/2 - B_t)$. Uno simplemente genera aleatoriamente $m(x) \in R(q/2 - B_m)$, hasta que $t(x) = h(x)m(x) \in R(q/2 - B_t)$.

J. Hoffstein et al., comentan en [11] que se puede concluir que la región de firmas contiene una gran fracción del lattice $L^{NT}(q/2, q/2)$ (al menos 30 % para los conjuntos de parámetros que consideramos) esto sucede después de un pequeño número de iteraciones en una transcripción de la forma:

$$[(m(x), t(x))_i, (m_p(x), t_p(x))_i], \quad i = 1, 2, \ldots,$$

donde $(m(x), t(x))_i$ se produce como lo dicho anteriormente y la igualdad de cada hash con su posible firma $(m_p(x), t_p(x))_i = (m(x), t(x))_i \mod p$, se distribuye uniformemente en el lattice $L^{NT}(q/2 - B_m, q/2 - B_t) \times [R(p/2) \times R(p/2)]$, por la suposición 1.

Por la proposición 6.6.4, y por la suposición de que un documento hash es uniforme en $R(p/2) \times R(p/2)$ se tiene una transcripción indistinguible de una transcripción producida por un firmante honesto. La única diferencia entre las dos transcripciones es que el firmante (falsificador) que usó solo la clave pública $h(x)$, solo conoce los mensajes (o documentos) $\mu_i(x)$, tales que $hash(h(x), \mu_i(x)) = (m_p(x), t_p(x))$.

6.7. Ejemplo de esquema de firma digital $NTRUSign$

En esta sección mostramos el funcionamiento del esquema de firma digital $NTRUSign$ con un ejemplo. Para ilustrar este esquema de firma, de manera que se enfatice la visualización de su mecanismo, tomaremos los parámetros pequeños y no como se establece en la tabla 6.1.

Generación de claves.

Seleccionamos los parámetros:

$$(n, p, q, d, k, m, B_k, B_m, B_t) = (6, 2, 23, 2, 3, 3, 11, 4, 2).$$

Seleccionamos polinomios:

$$f(x) = 2x^5 + 2x^4 + 2x^2 - 2x - 2.$$
$$g(x) = -x^4 + x^3 + x^2 + x - 1.$$

Los polinomios inversos de $f(x)$ en R_q y $g(x)$ en R_p son:

$$F_q(x) = 6x^5 + 10x^4 + 10x^3 + x^2 + 4x + 2.$$
$$g_p(x) = x^5 + x^4 + x^3 + x^2 + 1.$$

Calculamos la clave pública $h(x)$, la cual resulta ser:

$$h(x) = 11x^5 + 3x^4 + 3x^3 + 9x^2 + 15x + 19.$$

Verificamos la condición $NORMF(f(x))$ y $NORMF(g(x))$:

$$F1(x) = \sum_{i=0}^{5} f(x)x^i = 2x^5 - 2x^4 - 6x^3 - 6x^2 - 10x - 6.$$
$$G1(x) = \sum_{i=0}^{5} g(x)x^i = x^5 + x^4 + 3x^3 + x^2 - x - 3.$$

Entonces, claramente $NORMF(F1(x)) = 10 \leq B_k$ y $NORMF(G1(x)) = 3 \leq B_k$, con $B_k = 11$.

Firmaremos en mensaje: “La criptografía es divertida”.

Algoritmo de firma.

Aplicamos la función SHA-3-512 para calcular el resumen del mensaje.

$msg\_digest =$ 01100001001110000011011000110100001100110110001101100110001101010
11000010011000000110001011001100110000100110110001100100011000000
11100000110010001101010011001000110101001101000011001100111001001
10000011001100011000100111001001110000110010000110111011000010110
01010011011100110111011001010011100000110111011001100011001001100
01100111001001110000110010101100001001100000011001001100010001100
11001110000011100001100110011001100011010000110010011000110110010
00110011000110101011000110011010101100011011000110011100100110001
00110101001100010110001001100010001100100110010100110010011001010
01101010011100100111001011000010011011001100001001101010011010100
11011000110101001100110011001100110001011001100110001100111001001
10010011000110110000100110110001101110011001101100010011001000011
01000110001000111001011000110011000000111001011001000110000101100
01101100110011000110011100101100100011001000011011101100110011000
01001100000011100100110001011000010011100001100011011000100110011
00011011100110100011001010011011101100101001101111.

El hash del mensaje es:

$$m_p(x) = x^2.$$
$$t_p(x) = x^2 + x.$$

Seleccionamos un polinomio aleatorio en el conjunto $R(A)$,

$$r(x) = -x + 1.$$

Calculamos los siguientes polinomios:

$$m_0(x) = x^2 - 2x + 2.$$
$$t_0(x) = 8x^5 + 9x^4 + 3x^3 + 7x^2 + 3x + 2.$$
$$a(x) = x^5 + x^4.$$
$$m'(x) = -4x^5 - 4x^4 - 4x^3 - x^2 - 4x + 2.$$
$$t'(x) = 8x^5 + 8x^4 + 4x^3 + 7x^2 + x.$$
$$a(x)f(x) = 2x^{10} + 4x^9 + 2x^8 + 2x^7 - 4x^5 - 2x^4.$$
$$a(x)g(x) = -x^9 + 2x^7 + 2x^6 - x^4.$$

Las normas infinito de los siguientes polinomios son:

$$||a(x)f(x)||_\infty = 4, \qquad ||a(x)g(x)||_\infty = 2, \qquad ||m'(x)||_\infty = 8, \qquad ||t'(x)||_\infty = 4.$$

Vemos que las constantes $B_m = 4$, $B_t = 2$, $q/2 - B_m = 19/2$ y $q/2 - B_t = 21/2$ acotan a las cuatro normas anteriores, como lo establece el algoritmo de firma, es decir, se tiene que:

$$||a(x)f(x)||_\infty = 4 \leq B_m.$$
$$||a(x)g(x)||_\infty = 2 \leq B_t.$$
$$||m'(x)||_\infty = 8 \leq q/2 - B_t = 9{,}5.$$
$$||t'(x)||_\infty = 4 \leq q/2 - B_m = 7{,}5.$$

Por lo tanto. la firma para el mensaje “La criptografía es divertida”, es:

$$m'(x) = -4x^5 - 4x^4 - 4x^3 - x^2 - 4x + 2.$$
$$t'(x) = 8x^5 + 8x^4 + 4x^3 + 7x^2 + x.$$

Capítulo 7

Firma digital con ETRU

En este capítulo se expone el propósito central de la presente tesis, que es dar una propuesta del esquema de firma digital $NTRUSign$ sobre el anillo de los enteros de Eisenstein $\mathbb{Z}[\omega]$ inspirados en el esquema de firma digital $NTRUSign$ propuesto por J. Hoffstein et al. en [7] y [11] como una alternativa más para lograr firmar documentos digitales de manera segura; se introducen los conceptos básicos que se utilizarán para acoplarse al anillo de los enteros de Eisenstein, los cuales son similares a los vistos en el capítulo 5, y que son estudiados en [20]. Se proponen los algoritmos de firma digital y verificación digital, así como un análisis de seguridad. Finalmente se proporciona un ejemplo para realizar firmas digitales con este nuevo esquema.

7.1. Notación

En esta sección presentamos la notación necesaria para emplear la propuesta de esquema de firma digital NTRU sobre los enteros de Eisenstein, la cual facilitará la comprensión de todo el desarrollo y análisis de este esquema de firma.

El esquema que se definirá en la siguiente sección utiliza la norma euclidiana de $\mathbb{Z}[\omega]$, ya que trabajar con está norma tiene ventajas como lo indica la observación 5.3.2 del capítulo 5, es decir, que usar la norma euclidiana mejora el funcionamiento del algoritmo LLL, el cual se implementa en los ataques de lattices contra este esquema de firma.

Sea $a(x) = a_0 + a_1x + \cdots + a_nx^n$ un polinomio con coeficientes en $\mathbb{Z}[\omega]$, la ***norma infinito*** de $a(x)$ respecto a la norma euclidiana de $\mathbb{Z}[\omega]$ se define por,

$$||a(x)||_\infty = ||(a_0, a_1, \ldots, a_n)||_\infty = \max_i N(a_i),$$

donde $N(\cdot)$ es la norma euclidiana de $\mathbb{Z}[\omega]$.

La ***norma euclidiana*** del polinomio $a(x)$ es,

$$\bar{N}(a(x)) = \bar{N}((a_0, a_1, \ldots, a_n)) = \sqrt{N(a_0)^2 + N(a_1)^2 + \cdots + N(a_n)^2}.$$

La ***norma euclidiana de dos o más polinomios concatenados***, de la forma $a(x) = a_0 + a_1x + \cdots + a_nx^n$ y $b(x) = b_0 + b_1x + \cdots + b_nx^n$, con coeficientes en $\mathbb{Z}[\omega]$ se define por,

$$\bar{N}(a(x), b(x)) = \bar{N}(a_0, a_1, \ldots, a_n, b_0, b_1, \ldots, b_n) = \sqrt{\sum_{i=0}^{n} N(a_i)^2 + \sum_{i=0}^{n} N(b_i)^2}.$$

En analogía con el esquema de firma $NTRUSign$, utilizaremos uno de los anillos de polinomios, pero ahora con la reducción del ideal $\langle x^n - 1\rangle$, que se utilizó en la sección 6.5.3 del capítulo 6, es decir, nuestro anillo base será el anillo de polinomios $R_q = \mathbb{Z}_q[\omega][x]/\langle x^n - 1\rangle$, teniendo así, la notación $R = R_q$.

Sabemos que las unidades de $\mathbb{Z}[\omega]$ son $U(\mathbb{Z}[\omega]) = \left\{\pm 1, \pm\omega, \pm\omega^2\right\}$. Ahora, recordamos la definición del conjunto de polinomios ternarios en $\mathbb{Z}[\omega]$.

Definición 7.1.1. *Definimos el conjunto de polinomios ternarios en $\mathbb{Z}[\omega]$ de la siguiente manera:*

$$T(u_1, u_2) = \left\{ a(x) \in R : \begin{array}{c} a(x) \text{ tiene } u_1 \text{ coeficientes iguales a } 1,\ u_2 \text{ coeficientes} \\ \text{de cada una del resto de las unidades y el resto} \\ \text{de sus coeficientes iguales a } 0 \end{array} \right\}$$

Con la notación anterior establecemos el subconjunto $R(k)$ de R como sigue,

$$R(k) = \{f(x) \in R : ||f(x)||_\infty \leq N(k)\},$$

donde k es un entero de Eisenstein fijo.

Una matriz de rotación cíclica de un polinomio $f(x)$ sobre el anillo R es una matriz $M = (f_1, f_2, \ldots, f_n)^T$, con $f_i = f(x)x^{i-1} \mod (x^n - 1)$. Particularmente para referirnos a la norma infinito de elementos de M (vectores), establecemos la notación $NORMF$, así que $NORMF$ se calcula como $||t(x)||_\infty$, donde $t(x) = \sum_{i=0}^{n-1} f(x)x^i \mod (x^n - 1)$.

Notemos que si un usuario desea firmar algún documento electrónico en claro, con el esquema que presentaremos en la siguiente sección, basta que considere el documento convertido en su forma binaria, y para esto solo requiere de algún código estándar o convertidor de caracteres a números binarios, por ejemplo, el código ASCII es un código que se usa con mucha frecuencia para convertir documentos (caracteres) en claro a su forma binaria.

Por lo tanto, en este esquema de firma, los mensajes serán considerados como cadenas de bits y se representaran simplemente por μ o por μ_i, cuando es un mensaje específico.

También, consideraremos subconjuntos de lattices L_h que consisten en vectores de norma acotada y los dentaremos por,

$$L_h(k_1, k_2) = L_h \cap (R(k_1) \times R(k_2)),$$

donde L_h es un lattice, k_1 y k_2 son enteros de Eisenstein.

Dado un anillo R y $p \in R$, el conjunto,

$$pR = \{px : x \in R\},$$

lo llamaremos el ***conjunto de los múltiplos del elemento*** $p \in R$.

7.2. Definición de esquema de firma digital $NTRUSign$ sobre los enteros de Eisenstein

El siguiente esquema de firma digital está inspirado en el esquema de firma digital $NTRUSign$ del capítulo 6, el cual fue propuesto por J. Hoffstein et al. en [11]. Este esquema utiliza los anillos

base $R = R_q = \mathbb{Z}_q[\omega][x]/\langle x^n - 1\rangle$ y $R_p = \mathbb{Z}_p[\omega][x]/\langle x^n - 1\rangle$ y tiene un funcionamiento similar al esquema $NTRUSign$. Además, este nuevo esquema propuesto goza de algunas ventajas de seguridad en comparación con el esquema $NTRUSign$, las cuales se estudiarán más adelante.

Definimos el esquema de firma digital $NTRUSign$ sobre el anillo de los enteros de Eisenstein, mismo que denotaremos por $ETRUSign$, por la tupla:

$$(\mathcal{M}, R \times R, pR(p) \times R(p), S, V),$$

donde

a) $\mathcal{M}$ es el conjunto de mensajes posibles en su forma binaria.

b) $R \times R$ es el producto cartesiano del anillo R consigo mismo.

c) $pR(p) \times R(p)$, es el producto cartesiano del conjunto de polinomios $pR(p)$, con el conjunto de polinomios $R(p)$.

d) S es el conjunto de algoritmos de firma $fir_k : \mathcal{M} \to R \times R$.

e) V es el conjunto de algoritmos de verificación $ver_k : R \times R \times \mathcal{M} \to \{válida,\ no\ válida\}$.

Note que el algoritmo de firma entrega como firma digital a un par de polinomios $(m(x), t(x))$ para el mensaje μ, mientras que el algoritmo de verificación tiene por entrada a la terna $(m(x), t(x), \mu)$. El hecho de que el algoritmo de verificación tenga como entrada a dicha terna de polinomios en vez de un par de polinomios se debe a que puede haber distintos pares de polinomios, digamos $(m'(x), t'(x))$, que sean firma válida para μ.

El esquema de firma $ETRUSign$ se lleva a cabo de la siguiente manera:

7.2.1. Generación de claves

Enseguida mostramos el proceso para que algún usuario interesado en firmar un documento electrónico con $ETRUSign$ genere las claves adecuadas.

El usuario interesado en firmar un mensaje o documento electrónico deberá fijar los parámetros $(n, k, m, p = 2+\omega, q, d), B_k, B_m, B_t$, tales que n, k, m, d son enteros positivos con $n = k+m$, $p = 2+\omega$ y $q = a + b\,\omega$ son enteros de Eisenstein, donde $\nu(q) > \nu(p)$ y tal que $\nu(q)$ es un primo racional. Los elementos B_k, B_m, B_t son constantes de números eisenianos que se determinan de manera experimental.

En la sección 7.2.4 se dará una propuesta para fijar a los números eisenianos B_k, B_m, B_t.

La manera de establecer las claves de $ETRUSign$ es como sigue:

- Aldo selecciona de manera aleatoria polinomios $F(x) \in T(d+1, d)$, $g(x) \in T(d+1, d)$, de manera que el polinomio $x-1$ no sea factor de $F(x)$ ni de $g(x)$, para así, reducir la posibilidad de que estos polinomios no sean invertibles. Aldo puede usar el algoritmo 11 de la sección 6.5.2 para generar a dichos polinomios, luego, se establece $f(x) = pF(x)$.

- Enseguida calcula los polinomios $g^{-1}(x) \in R_p$ y $F_q(x) = f^{-1}(x) \in R_q$. En caso de que $g(x)$ y $f(x)$ no sean invertibles, se deben elegir nuevos polinomios $g(x)$ y $f(x)$ que sí tengan inverso multiplicativo en R_p y R_q respectivamente.

- Se verifica la condición de norma para el polinomio $f(x)$, $NORMF(f(x)) < N(B_k)$ y la condición de norma para $g(x)$, $NORMF(g(x)) < N(B_k)$. En caso de que alguna de las dos condiciones no se cumpla, Aldo deberá elegir nuevos polinomios $f(x)$ o $g(x)$ que sí cumplan lo anterior.

- Finalmente se calcula el producto $h(x) = g(x)F_q(x) \mod q$.

Los parámetros y claves que ha generado Aldo para el esquema *ETRUSign* quedan de la siguiente manera:

1. Los parámetros n, k, m, p, q, B_m, B_t son públicos.

2. Se deja como clave pública al polinomio $h(x)$.

3. La clave privada serán la pareja de polinomios $(f(x), g(x))$.

7.2.2. Algoritmo de firma digital para $ETRUSign$

El siguiente algoritmo calcula la firma digital *ETRUSign* para un mensaje μ, utilizando una clave privada $(f(x), g(x))$.

Supongamos que Aldo desea firmar un mensaje o documento digital μ. Entonces debe considerar los polinomios seleccionados de la sección anterior, los cuales son: $h(x), f(x), g(x)$. Debe considerar una sucesión aleatoria de bits, s, como fuente de aleatoriedad. Con ellos genera una firma $(m'(x), t'(x))$ para el mensaje μ.

Entrada: $(f(x), g(x), h(x), \mu, s)$, donde $(f(x), g(x))$ es una clave privada, $h(x)$ su correspondiente clave pública, μ es el mensaje a firmar, que está previamente convertido en forma binaria y s es una cadena aleatoria de bits.

1. Aldo genera el resumen del mensaje mediante una función hash aprobada por el Instituto Nacional de Estándares y Tecnología (NIST). Por ejemplo, Aldo puede usar la función SHA-3-512 como función hash, y así obtener el resumen del mensaje *msg_digest*.

2. **Repetir**:

   2.1 Aldo elige dos polinomios aleatorios uniformes. Por ejemplo, utilizando el algoritmo 12 de la sección 6.5.2, con parámetros k y m, módulo $p = 2 + \omega$, y cadena aleatoria al resumen del mensaje *msg_digest* para obtener polinomios $m_p(x) \in R(p)$ y $t_p(x) \in R(p)$. Note que los polinomios $m_p(x)$ y $t_p(x)$ tienen exactamente un total de k coeficientes y t coeficientes, respectivamente.

   2.2 Aldo elige un polinomio aleatorio uniforme de la misma manera como en el paso 2.1, con parámetros n, módulo $A_E = \frac{\sqrt{2}N(p)}{3}[\nu(q) + N(p)]$, y semilla a la sucesión aleatoria s. De este modo se obtiene un polinomio $r(x) \in R(A_E)$.

   2.3 Aldo calcula los siguientes polinomios:

   a) $m_0(x) = m_p(x) + pr(x)$.
   b) $t_0(x) = h(x)m_0(x) \mod q, \quad \text{con } t_0(x) \in H_q$.
   c) $a(x) = g^{-1}(x)(t_p(x) - t_0(x)) \mod p, \quad \text{con } a(x) \in H_p$.
   d) $m'(x) = m_0(x) + a(x)f(x)$.
   e) $t'(x) = t_0(x) + a(x)g(x)$,

   donde H_q y H_p son los dominios fundamentales de los ideales (q) y (p), respectivamente.

3. **Hasta que:**

$$||a(x)f(x)||_\infty \leq N(B_m), \;\; ||a(x)g(x)||_\infty \leq N(B_t),$$
$$||m'(x)||_\infty \leq \nu(q) - N(B_m), \;\; ||t'(x)||_\infty \leq \nu(q) - N(B_t).$$

4. **Salida:** $(m'(x), t'(x), \mu)$.

7.2.3. Algoritmo de verificación

La firma $(m'(x), t'(x))$ para el mensaje μ, obtenida por el algoritmo anterior se verifica de la siguiente manera:

1. **Entrada:** $(m'(x), t'(x), \mu, h(x), s)$,
2. El usuario que desee verificar una firma debe elegir dos polinomios aleatorios uniformes. Por ejemplo, mediante el algoritmo 12 de la sección 6.5.2, con parámetros k y m, módulo $p = 2+\omega$, y cadena aleatoria al resumen del mensaje *msg_digest*, y así, obtener polinomios $m_p(x) \in R(p)$ y $t_p(x) \in R(p)$.

   Luego, verifica las siguientes condiciones,
3. Si $t'(x) \not\equiv h(x)m'(x)$ (mód q), envíe "no válida"si es así.
4. Si $||m'(x)||_\infty > \nu(q) - N(B_m)$ o $||t'(x)||_\infty > \nu(q) - N(B_t)$, envíe "no válida"si es así.
5. Si $(m'(x), t'(x)) \not\equiv (m_p(x), t_p(x))$ (mód p), envíe "no válida"si es así.
6. **Salida:** "Válida".

7.2.4. Parámetros

A continuación se presenta una tabla de valores específicos para los parámetros del esquema de firma $ETRUSign$, que hacen a este esquema resistente a los ataques de falsificaciones y a los ataques de transcripciones.

Parámetros	R	n	q	d	B_k, B_m, B_t
$ETRUSign$	$\frac{\mathbb{Z}_q[x][\omega]}{(x^n-1)}$	1024	$249 + 256\omega$	52	$29 + 31\omega$, $53 + 59\omega$, $29 + 31\omega$

Tabla 7.1: Parámetros para el esquema de firma $ETRUSign$.

7.3. Análisis de seguridad

El esquema de firma digital $ETRUSign$ como el esquema de firma $NTRUSign$ son esquemas basados en anillos de polinomios con coeficientes en $\mathbb{Z}[\omega]$ y $\mathbb{Z}$, respectivamente y parte de su seguridad se basa en lattices que son generados por los coeficientes del polinomio de la clave pública $h(x)$, en ambos esquemas; por lo que, la seguridad está determinada por la dificultad de encontrar el único vector más corto en el lattice ETRU y NTRU, respectivamente.

7.3.1. Ataque con solo la clave pública $h(x)$

Implementamos un ataque similar al de las secciones 6.6.1 y 5.3, considerando que cualquier usuario puede tener acceso a la clave pública $h(x)$. Se puede emprender un ataque de lattices y aplicar técnicas de reducción de lattices, como el algoritmo LLL, para intentar recuperar la clave privada $(f(x), g(x))$.

El lattice generado por la clave pública $h(x)$ del esquema de firma $ETRUSign$ lo denotaremos por L^{ET}. Siguiendo la analogía del ataque de lattices de la sección 5.3 del capítulo 5, recordamos los siguientes conceptos,

Si $\alpha = a + b\omega \in \mathbb{Z}[\omega]$, definimos,

$$\langle \alpha \rangle = \begin{bmatrix} a & b \\ -b & a-b \end{bmatrix}.$$

Vemos que si $\beta = c + d\omega \in \mathbb{Z}[\omega]$, el producto,

$$[c, d] \begin{bmatrix} a & b \\ -b & a-b \end{bmatrix} = [ac - bd, bc + ad - bd], \tag{7.1}$$

es la representación del producto de los enteros eisenianos α y β en la base $\{1, \omega\}$.

Consideramos la matriz B^{ET}, que como sabemos de la sección 5.3 del capítulo 5, es base del lattice L^{ET}, la cual está dada por,

$$B^{ET} = \begin{bmatrix} \lambda \langle I \rangle & \langle H \rangle \\ 0 & \langle qI \rangle \end{bmatrix},$$

donde $\langle H \rangle$ es la matriz circulante por bloques, $\langle h_i \rangle$, de la clave pública $h(x)$, λ es la constante de equilibrio, $\langle I \rangle$ es una matriz identidad de tamaño $2n \times 2n$ de bloques $\langle I_{2\times 2} \rangle$, y q es el parámetro que se selecciona en la sección 7.2.1.

Notemos que para cada vector $[f', u] \in \mathbb{Z}^{4n}$, donde f' y u representan los coeficientes de los polinomios $f'(x)$ y $u(x)$, en la base $\{1, \omega\}$ como en (7.1). Con esta observación es fácil ver que los vectores f' y u están en $\mathbb{Z}^{2n}$. Luego, tenemos que,

$$[f'\ u] B^{ET} = [f'\ u] \begin{bmatrix} \lambda \langle I \rangle & \langle H \rangle \\ 0 & \langle qI \rangle \end{bmatrix} = [\lambda f', g].$$

El desarrollo explícito de la ecuación anterior es el mismo que el de la sección 5.3 del capítulo 5, por esa razón omitimos en esta parte ese desarrollo.

Por lo tanto, el vector $[\lambda f', g]$ está contenido en el lattice L^{ET}. Este vector es relativamente corto, por lo que, podemos usar técnicas de reducción de bases para intentar recuperar el vector $[\lambda f, g]$, y así recuperar la clave privada del esquema de firma $ETRUSign$.

Es importante notar que por la observación 5.3.2 del capítulo 5, se considera el uso de las normas euclidianas para el cálculo de las normas de los polinomios $f(x)$ y $g(x)$, y así, brindar un cálculo adecuado del valor τ.

Si τ_c representa la longitud del vector objetivo $[\lambda f, g]$, entonces,

$$\tau_c = \sqrt{\lambda^2 N(f(x))^2 + N(g(x))^2},$$

donde $N(f(x))$ es la norma euclidiana de $f(x)$.

Dado que $f(x) = pF(x)$, donde el polinomio $F(x)$ se selecciona en $T(d+1, d)$ como en la sección 7.2.1, entonces $F(x)$ tiene $d+1$ coeficientes 1, d coeficientes de cada del resto de las unidades y el resto de sus coeficientes iguales a cero y de manera que el polinomio $x-1$ no sea factor de $f(x)$. Entonces, la forma más general que puede tomar polinomio $f(x)$ es:

$$\begin{aligned} f(x) = p[&1 + x + \cdots + x^u + (-1)x^{u+1} + (-1)x^{u+2} + \cdots + (-1)x^{2u} + \\ &+ \omega x^{2u+1} + \omega x^{2u+2} + \cdots + \omega x^{3u} + (-\omega)x^{3u+1} + (-\omega)x^{3u+2} + \cdots + (-\omega)x^{4d} + \\ &+ \omega^2 x^{4u+1} + \omega^2 x^{4u+2} + \cdots + \omega^2 x^{5u} + (-\omega)^2 x^{5u+1} + (-\omega)^2 x^{5u+2} + \cdots + (-\omega)^2 x^{6u}]. \end{aligned}$$

Consideramos los siguientes productos de números eisenianos, los cuales son los productos que se involucran en el polinomio $f(x)$, y que se utilizarán para simplificar los cálculos de la forma $p \cdot u$, con $u \in U(\mathbb{Z}[\omega])$, en el cálculo de $N(f(x))$. Si $p = 2 + \omega$ y $u \in U(\mathbb{Z}[\omega])$, entonces,

$$\begin{aligned} p \cdot 1 &= 2 + \omega. \\ p \cdot -1 &= -2 - \omega. \\ p \cdot \omega &= (2+\omega)\omega = 2\omega + \omega^2 = \omega - 1. \\ p \cdot -\omega &= (2+\omega) - \omega = -2\omega - \omega^2 = -\omega + 1. \\ p \cdot \omega^2 &= (2+\omega)\omega^2 = 2\omega^2 + \omega^3 = -2\omega - 1. \\ p \cdot -\omega^2 &= (2+\omega) - \omega^2 = -2\omega^2 - \omega^3 = 2\omega + 1. \end{aligned}$$

Procedemos a calcular la norma del polinomio $f(x)$ y $g(x)$,

$$\begin{aligned} N(f(x))^2 = &\underbrace{N(2+\omega)^2 + \cdots + N(2+\omega)^2}_{u+1 \text{ veces}} + \underbrace{N(-2-\omega)^2 + \cdots + N(-2-\omega)^2}_{d \text{ veces}} + \\ &+ \underbrace{N(\omega-1)^2 + \cdots + N(\omega-1)^2}_{d \text{ veces}} + \underbrace{N(-\omega+1)^2 + \cdots + N(-\omega+1)^2}_{d \text{ veces}} + \\ &+ \underbrace{N(-2\omega-1)^2 + \cdots + N(-2\omega-1)^2}_{d \text{ veces}} + \underbrace{N(2\omega+1)^2 + \cdots + N(2\omega+1)^2}_{d \text{ veces}} \\ &= (d+1)N(2+\omega)^2 + dN(-2-\omega)^2 + dN(\omega-1)^2 + \\ &+ dN(-\omega+1)^2 + dN(-2\omega-1)^2 + dN(2\omega+1)^2 \\ &= (d+1)5 + d5 + d2 + d2 + d5 + d5 \\ &= 24d + 5. \end{aligned}$$

Entonces $N(f(x)) = \sqrt{24d+5}$.

De manera similar podemos realizar los mismos cálculos anteriores para obtener la norma euclidiana del polinomio $g(x)$, el cual se elige con las mismas características que el polinomio $f(x)$. La forma más general de $g(x)$ es:

$$\begin{aligned} g(x) = &1 + x + \cdots + x^d + (-1)x^{d+1} + (-1)x^{d+2} + \cdots + (-1)x^{2d} + \\ &+ \omega x^{2d+1} + \omega x^{2d+2} + \cdots + \omega x^{3d} + (-\omega)x^{3d+1} + (-\omega)x^{3d+2} + \cdots + (-\omega)x^{4d} + \\ &+ \omega^2 x^{4d+1} + \omega^2 x^{4d+2} + \cdots + \omega^2 x^{5d} + (-\omega)^2 x^{5d+1} + (-\omega)^2 x^{5d+2} + \cdots + (-\omega)^2 x^{6d}. \end{aligned}$$

Notemos que en este caso la forma de $g(x)$ coincide con la del polinomio $f(x)$ de la sección 5.3 del capítulo 5. Por lo tanto, $g(x)$ tendrá norma euclidiana igual a la calculada para el polinomio $f(x)$ en la sección ya mencionada, así que,

$$N(g(x))^2 = 8d + 1,$$

es decir, $N(g(x)) = \sqrt{8d+1}$.

Ahora, sustituimos los valores de $N(f(x))$ y $N(g(x))$ en el valor de la longitud del vector objetivo, $(\lambda f, g)$,

$$\begin{aligned}\tau_c &= \sqrt{\lambda^2 N(f(x))^2 + N(g(x))^2} \\ &= \sqrt{\lambda^2[24d+5] + 8d + 1}.\end{aligned}$$

El vector $(\lambda f, g)$ es un vector relativamente corto en el lattice L^{ET}, por lo que se puede emplear el algoritmo LLL para tratar de encontrar un vector de longitud τ_c y con ello la clave secreta $(f(x), g(x))$.

7.3.2. Elección de la constante de equilibrio

Análogamente a la sección 5.3.1 del capítulo 5, buscamos que la constante λ se elija de forma que maximice la eficiencia de encontrar vectores cortos en el lattice L^{ET}. Entonces, podemos maximizar la relación s/τ_c, donde s es la longitud del vector esperado más corto en L^{ET} y τ_c la longitud objetivo.

Obtenemos el valor de la longitud s mediante la fórmula de la heurística gaussiana 1, vista en la sección 6.6.2 del capítulo 6 y en la sección 2.5 del capítulo 2.

De la sección 5.3.1 del capítulo 5, podemos ver que se tiene el mismo lattice L^{ET} que se usó en el ataque de lattices al criptosistema ETRU, por lo que, el valor de s es el mismo que el de está sección, es decir, la longitud esperada del vector distinto de cero más corto es,

$$s = \sqrt{\frac{2n\lambda||q||_2}{\pi e}}.$$

De este modo, la relación s/τ_c queda de la siguiente manera,

$$\begin{aligned}s/\tau_c &= \sqrt{\frac{\frac{2n\lambda||q||_2}{\pi e}}{\lambda^2[24d+5] + 8d + 1}} \\ &= \sqrt{\frac{2n\lambda||q||_2}{\pi e\lambda^2[24d+5] + \pi e(8d+1)}} \\ &= \sqrt{\frac{2n\lambda||q||_2}{\pi e(\lambda^2 N(f(x))^2 + N(g(x))^2)}}.\end{aligned}$$

Así, maximizar la razón s/τ_c se reduce a maximizar el término,

$$\frac{\lambda}{\lambda^2 N(f(x))2 + N(g(x))^2},$$

que es el mismo término obtenido al final de la sección 5.3 del capítulo 5, en el cual se dan los mismo argumentos de la sección 3.4.1 del capítulo 3 para poder considerar $\lambda = 1$, así como los argumentos necesarios para el cual el vector objetivo $(\lambda f, g)$ es adecuado.

Notemos que si consideramos $\lambda = 1$, podemos justificar con un argumento similar al de la sección 3.4.1 del capítulo 3 que $f(x) \approx g(x)$. Este hecho se considerará en la siguiente sección, cuando se calcule el valor de τ (longitud objetivo) de manera explícita.

Observación 7.3.1. *Hemos calculado un valor explícito de la longitud objetivo τ_c dado por $\tau_c = \sqrt{\lambda^2[21d+5]+8d+1}$, mientras que el cálculo de τ en el ataque de lattices al criptosistema ETRU en la sección 5.3 del capítulo 5, fue de $\tau_c = \sqrt{\lambda^2[8d+1]+8d}$, a pesar de que se realizaron los mismos procedimientos, se tienen cambios significativos, esto se debe a la forma que pueden tomar los polinomios $f(x)$ y $g(x)$ en el criptosistema ETRU y en el esquema de firma ETRUSign. Esto contrasta con el hecho de que la longitud del vector $(\lambda f, g)$ en el criptosistema ETRU es mejor (o más pequeña) que en el esquema ETRUSign, dando así, una desventaja para algún usuario que realice un criptoanálisis con este tipo de ataque.*

7.3.3. Ataque de falsificaciones

En esta sección estudiamos un ataque de falsificación de firmas similar al de la sección 6.6.2 del capítulo 6, ajustado a los lattices L^{ET} y $p\mathbb{Z}[\omega]^{4n}$, el cual esta basado en reducción de bases para lattices. Este ataque se puede implementar para intentar obtener posibles firmas válidas para este esquema de firma.

Recordemos que la falsificación de una firma se puede lograr si se puede resolver el problema de vector más corto aproximado asociado a la intersección de los lattices L^{ET} y $p\mathbb{Z}[\omega]^{4n}$. Por lo tanto, la tarea de falsificación se puede resolver encontrando un vector en $L^{ET} \cap p\mathbb{Z}[\omega]^{4n}$ que cumpla con los requisitos de congruencia módulo p del algoritmo de verificación de firma, y que esté lo suficientemente cerca del lattice de intersección para satisfacer los requisitos de longitud (norma).

Proposición 7.3.1. *Sea r un entero racional par, mayor o igual a 2. Sean $L_1 \subset \mathbb{Z}[\omega]^r$ y $L_2 \subset \mathbb{Z}[\omega]^r$ lattices de rango r, suponga que $t_1, t_2 \in \mathbb{Z}[\omega]^r$ son vectores arbitrarios y sea $M = (L_1 + t_1) \cap (L_2 + t_2)$ la intersección de las traslaciones de L_1 y L_2. Hacemos las siguientes suposiciones:*

1. $\mathrm{mcd}(\det(L_1), \det(L_2)) = 1$.
2. *$t_1 \notin L_1$ o $t_2 \notin L_2$, por lo que, en particular $M \neq L_1 \cap L_2$.*

Entonces se cumplen las siguientes condiciones:

a) $\det(L_1 \cap L_2) = \det(L_1)\det(L_2)$.

b) $M \neq \emptyset$.

c) Para cada $w_0 \in M$, la función:

$$L_1 \cap L_2 \to M, \qquad v \mapsto v + w_0$$

es una biyección.

d) Sea $w_0 \in M$ y suponga que $w' \in M$ es un vector distinto de cero más corto en M. Entonces $w_0 - w'$ resuelve el problema del vector más cercano en $L_1 \cap L_2$ para el vector w_0 (Esto es cierto para cualquier norma en $\mathbb{Z}^r$), por lo que, en particular es cierto para la norma euclidiana y la norma infinito).

Demostración. La prueba de los inciso $a), c), d)$ son similares a las pruebas de estos mismos incisos de la proposición 6.6.1 del capítulo 6.

Para la prueba del inciso $b)$ tenemos lo siguiente:

Hacemos $D_i = \det(L_i)$ para $i = 1, 2$. Probaremos que si $L \subset \mathbb{Z}[\omega]^r$ es un lattice con determinante D, entonces $D\mathbb{Z}[\omega]^r \subset L$.

Con un cálculo similar al que se vio en la sección 5.3.1 del capítulo 5, de base de un lattice, se puede verificar que la matriz,

$$B = \begin{bmatrix} \langle\alpha\rangle & 0 & \cdots & 0 \\ 0 & \langle\alpha\rangle & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \langle\alpha\rangle \end{bmatrix}_{r/2 \times r/2}$$

es base del lattice $\mathbb{Z}[\omega]^r$, donde $\alpha = 1 + \omega$.

Observe que el determinante de la matriz B es un número entero racional, pues $\langle\alpha\rangle$ es una matriz cuyas entradas son enteros racionales.

Entonces, también se puede verificar que la matriz,

$$DB = D \begin{bmatrix} \langle\alpha\rangle & 0 & \cdots & 0 \\ 0 & \langle\alpha\rangle & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \langle\alpha\rangle \end{bmatrix}_{r/2 \times r/2}$$

es una base para el lattice $D\mathbb{Z}[\omega]^r$.

Dado que la base DB es un múltiplo de la base B, claramente cualquier vector que se escriba como combinación lineal de la base DB se podrá escribir como combinación lineal de la base B.

Por hipótesis $L \subset \mathbb{Z}[\omega]^r$ y considerando que los vectores de B son un conjunto mínimo de generadores de $\mathbb{Z}[\omega]^r$, y como L es de rango r, se sigue que B también es un conjunto mínimo de generadores de L. Entonces si $v \in D\mathbb{Z}[\omega]^r$, v se puede escribir como combinación lineal de los vectores de B, que son base de $\mathbb{Z}[\omega]$, y que son generadores de L, es decir, si $v \in D\mathbb{Z}[\omega]^r$, entonces $v \in L$.

Por lo tanto, se ha probado que $D\mathbb{Z}[\omega]^r \subset L$.

La hipótesis de que $\text{mcd}(D_1, D_2) = 1$, significa que podemos encontrar $x, y \in \mathbb{Z}$, tal que,

$$xD_1 + yD_2 = 1.$$

Establecemos:

$$e_1 = yD_2 = 1 - xD_1, \qquad e_2 = xD_1 = 1 - yD_2,$$

y consideramos el siguiente vector $t = e_1t_1 + e_2t_2$, para tener que,

$$t - t_1 = (e_1 - 1)t_1 + e_2(t_2) = -xD_1t_1 + xD_1t_2 \in D_1\mathbb{Z}^r \subset L_1,$$

y similarmente,

$$t - t_2 = e_1t_1 + (e_2 - 1)t_2 = yD_2t_1 - yD_2t_2 \in D_2\mathbb{Z}^r \subset L_2.$$

Como el vector t está en M, entonces $M \neq \emptyset$. □

En este capítulo también haremos uso de las dos cantidades que están asociadas a los problemas de lattices, es decir la heurística 1, (6.2), y la heurística 2, (6.3), pero ahora con la observación siguiente.

Observación 7.3.2. *Notemos que la heurística 2 se expresa con lattices $L \subset \mathbb{Z}^n$ y respecto a la norma euclidiana usual de $\mathbb{R}^n$. En nuestros cálculos se considerará está misma heurística para lattices $L \subset \mathbb{Z}[\omega]^n$, con la norma euclidiana de $\mathbb{Z}[\omega]^n$.*

Consideremos el problema de falsificar una firma. El falsificador necesita encontrar un vector $(m', t') \in L^{ET}$, que satisfaga lo siguiente,

$$(m'(x), t'(x)) \equiv (m_p(x), t_p(x)) \quad \text{mód } p \quad \ldots\ldots \text{ condición de congruencia.} \tag{7.2}$$

$$||m'(x)||_\infty \leq \nu(q) - N(B_m) \quad \ldots\ldots \text{ condición de norma.} \tag{7.3}$$

$$||t'(x)||_\infty \leq \nu(q) - N(B_t) \quad \ldots\ldots \text{ condición de norma.} \tag{7.4}$$

Los polinomios $m_p(x), t_p(x)$ están en $R(p)$, entonces la condición de congruencia anterior puede formularse diciendo que la condición objetivo (m', t') está en la traslación del lattice $p\mathbb{Z}[\omega]^{4n}$ por el vector (m_p, t_p), en donde consideramos que $m_p = ([a_0, b_0], \ldots, [a_{n-1}, b_{n-1}])$, y $t_p = ([c_0, d_0], \ldots, [c_{n-1}, d_{n-1}])$, es decir, las entradas de los vectores m_p y t_p son números eisenianos, $m_i = a_i + b_i\omega$ y $t_i = c_i + d_i\omega$, respectivamente. Por lo tanto, el falsificador está buscando un vector corto en relación con la norma infinito en la intersección siguiente,

$$(m, t) \in L^{ET} \cap (p\mathbb{Z}^{4n} + (m_p, t_p)).$$

En la sección 5.3.1 del capítulo 5, se obtuvo que $det(L^{ET}) = \lambda^{2n}\nu(q)^{2n}$. Procedemos a calcular el determinante de $p\mathbb{Z}[\omega]^{4n}$.

Dado que $\mathbb{Z}[\omega]^{4n}$ es un lattice, podemos dar una base para este lattice. Sabemos que $\{1, \omega\}$ es base de $\mathbb{Z}[\omega]$, es decir, cualquier elemento $z \in \mathbb{Z}[\omega]$ se puede escribir como combinación lineal de 1 y ω. Definimos $\alpha = 1+\omega$ y probaremos que la siguiente matriz es una base para el lattice $\mathbb{Z}[\omega]^{4n}$.

Definimos a la matriz B, de la siguiente manera,

$$B = \begin{bmatrix} \langle\alpha\rangle & 0 & \cdots & 0 \\ 0 & \langle\alpha\rangle & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \langle\alpha\rangle \end{bmatrix}_{2n\times 2n}$$

Verificaremos que los vectores columna de la matriz B son linealmente independientes. Sean $z_1, z_2, \ldots, z_{2n} \in \mathbb{Z}[\omega]$, con $z_i = a_i + b_i\omega$. Si $\mathbf{0}$ representa el cero de $\mathbb{Z}[\omega]^{4n}$ con cada componente en su forma matricial, se puede expresar en su forma vectorial por $\mathbf{0} = [\langle 0\rangle, \ldots, \langle 0\rangle]^T_{2n\times 1}$, luego, vemos que,

$$\mathbf{0} = \langle z_1\rangle \begin{bmatrix} \langle\alpha\rangle \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \langle z_2\rangle \begin{bmatrix} 0 \\ \langle\alpha\rangle \\ \vdots \\ 0 \end{bmatrix} + \cdots + \langle z_{2n}\rangle \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \langle\alpha\rangle \end{bmatrix}. \tag{7.5}$$

Esto genera el sistema de ecuaciones,

$$\begin{aligned} \langle z_1 \rangle \langle \alpha \rangle &= \langle 0 \rangle \\ \langle z_2 \rangle \langle \alpha \rangle &= \langle 0 \rangle \\ \vdots \qquad &\quad \vdots \\ \langle z_{2n} \rangle \langle \alpha \rangle &= \langle 0 \rangle . \end{aligned}$$

Dado que $\langle \alpha \rangle$ es una matriz, con coordenadas enteras y como $det(\langle \alpha \rangle) \neq 0$, entonces existe $\langle \alpha \rangle^{-1}$ con coordenadas enteras, por lo tanto, podemos multiplicar por $\langle \alpha \rangle^{-1}$ en el sistemas de ecuaciones anterior para obtener que,

$$\begin{aligned} \langle z_1 \rangle &= \langle 0 \rangle \\ \langle z_2 \rangle &= \langle 0 \rangle \\ \vdots \quad &\quad \vdots \\ \langle z_{2n} \rangle &= \langle 0 \rangle , \end{aligned}$$

de donde, se sigue que $z_i = a_i + b_i\omega = 0 + 0\omega$, para $i = 1, 2, \ldots, 2n$. Por lo tanto, la ecuación (7.5) se cumple siempre que $\langle z_i \rangle = \langle 0 \rangle$, es decir, si cada número eiseniano z_i es cero. Se sigue que las columnas de la matriz B son linealmente independientes, y como hay un número finito de columnas linealmente independientes, podemos concluir que los vectores columna de la matriz B forman una base para el lattice $\mathbb{Z}[\omega]^{4n}$.

Notemos que, en la sección 7.2.1 para la generación de claves del esquema $ETRUSign$, hemos asignado el parámetro fijo $p = 2 + \omega$ y con la definición de $\alpha = 1 + \omega$, aplicamos un análisis similar al anterior para ver que la matriz,

$$pB = \begin{bmatrix} \langle p \rangle \langle \alpha \rangle & 0 & \cdots & 0 \\ 0 & \langle p \rangle \langle \alpha \rangle & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \langle p \rangle \langle \alpha \rangle \end{bmatrix}_{2n \times 2n}$$

es una base para el lattice $p\mathbb{Z}[\omega]^{4n}$.

Luego, tenemos que,

$$\det(p\mathbb{Z}[\omega]^{4n}) = \det(pB) = \prod_{i=1}^{2n} \det(\langle p \rangle \langle \alpha \rangle) = \det(\langle p \rangle \langle \alpha \rangle)^{2n}.$$

Como,

$$det(\langle p \rangle \langle \alpha \rangle) = det\left(\begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} \right) = det\left(\begin{bmatrix} 1 & 2 \\ -2 & -1 \end{bmatrix} \right) = -1 + 4 = 3,$$

se sigue que,

$$det(p\mathbb{Z}[\omega]^{4n}) = det(\langle p \rangle \langle \alpha \rangle)^{2n} = 3^{2n}.$$

Por lo tanto, tenemos que $\det(L^{ET}) = \lambda^{2n}\nu(q)^n$ y $\det(p\mathbb{Z}[\omega]^{4n}) = 3^{2n}$.

Una manera fácil para determinar cuándo los determinantes $\det(L^{ET})$ y $\det(p\mathbb{Z}[\omega]^{4n})$ son primos relativos es ver que las bases de las potencias son primos racionales.

Para $\det(L^{ET}) = \lambda^{2n}(\nu(q))^n$, consideramos que en la práctica se puede elegir la constante de equilibrio $\lambda = 1$ y así $\lambda^{2n} = 1$, y por la selección del parámetro q en la sección 7.2.1, la cual dice que $q = a + b\omega$ se elige de manera que $\nu(q)$ es un primo racional, entonces se tiene que $\nu(q)^n = r^n$ con r un primo racional. Además, como $\det(p\mathbb{Z}[\omega]^{4n}) = 3^{2n}$, entonces podemos concluir que $\det(L^{ET})$ y $\det(p\mathbb{Z}[\omega]^{4n})$ son primos relativos, por lo que, podemos usar la proposición 7.3.1 inciso a) para concluir que,

$$det(L^{ET} \cap p\mathbb{Z}[\omega]^{4n}) = \nu(q)^{2n}3^{2n}.$$

Entonces, la proposición 7.3.1, inciso d) nos dice que encontrar un vector corto en el lattice $L^{ET} \cap \ (p\mathbb{Z}[\omega]^{4n} + (m_p, t_p))$ es equivalente a resolver el problema τ-aprox-CVP en el lattice $L^{ET} \cap \ p\mathbb{Z}[\omega]^{4n}$. Un cálculo rutinario de la heurística gaussiana para el lattice $L^{ET} \cap p\mathbb{Z}[\omega]^{2n}$, nos permite ver que,

$$\gamma(L^{ET} \cap p\mathbb{Z}[\omega]^{4n}) = \sqrt{\frac{4n}{2\pi e}}(3^{2n}\nu(q)^{2n})^{1/4n} = \sqrt{\frac{2n}{\pi e}}3^{1/2}\nu(q)^{1/2} = \sqrt{\frac{6n}{\pi e}}||q||_2.$$

Solo queda estimar la longitud objetivo.

Veamos que $N(q) \leq \nu(q)$, siempre que $q \neq \pm\omega^2$, donde $q = a + b\omega$, $N(q)$ es la norma euclidiana aplicada a q y $\nu(q)$ es la función ν aplicada a q.

Sea $q = a+b\omega \in \mathbb{Z}[\omega]$, tal que $q \neq \pm\omega^2$. Si a y b tienen signos opuestos, entonces claramente el término $-ab$ es positivo y se satisface que,

$$\sqrt{a^2 + b^2} \leq a^2 + b^2 \leq a^2 + b^2 - ab,$$

es decir, se cumple que $N(q) \leq \nu(q)$.

Ahora, si a y b son positivos o a y b son negativos, entonces tenemos lo siguiente:

Caso 1. Supongamos que $a = b$, es decir, $q = a + a\omega$. Entonces,

$$N(q) = \sqrt{a^2 + a^2} = \sqrt{2a^2}, \qquad y \qquad \nu(q) = a^2 + b^2 - ab = a^2 + b^2 - a^2 = a^2.$$

Queremos que se cumpla que $\sqrt{2a^2} \leq a^2$, lo cual ocurre siempre que $a \neq 1$. Por hipótesis $q \neq \pm\omega^2$, es decir, $q \neq \pm(1 + \omega)$. Esto quiere decir que el coeficiente a es distinto de ± 1.

Por lo tanto, en este caso, se cumple que $N(q) \leq \nu(q)$, pues a es distinto de ± 1.

Caso 2. Supongamos que $a \neq b$.

Como a y b son enteros racionales y $a \neq b$, el término $(a - b)^2 \geq 1$, luego,

$$a^2 + b^2 \leq (a^2 + b^2)(a - b)^2 + a^2b^2. \tag{7.6}$$

Por otro lado, vemos que,

$$\begin{aligned}(a^2 + b^2 - ab)^2 &= a^4 + b^4 + a^2b^2 + 2a^2b^2 - 2a^3b - 2ab^3 \\ &= a^4 + b^4 + 3a^2b^2 - 2ab(a^2 + b^2) \\ &= (a^2 + b^2)^2 + a^2b^2 - 2ab(a^2 + b^2) \\ &= (a^2 + b^2)[a^2 + b^2 - 2ab] + a^2b^2 \\ &= (a^2 + b^2)(a - b)^2 + a^2b^2.\end{aligned}$$

Entonces, al sustituir el término $(a^2+b^2-ab)^2 = (a^2+b^2)(a-b)^2 + a^2b^2$ en la desigualdad (7.6), se sigue que,

$$a^2+b^2 \leq (a^2+b^2)(a-b)^2 + a^2b^2 = (a^2+b^2-ab)^2. \tag{7.7}$$

Finalmente sacamos raíz cuadrada a la desigualdad (7.7) para tener que $N(q) \leq \nu(q)$.

El criterio de rechazo en el algoritmo de firma dice que una firma válida $(m'(x), t'(x))$ tiene una norma superior como máximo $\nu(q) - N(\text{mín}(B_m, B_t))$. Por lo tanto, en particular, una firma válida satisface el límite de la norma euclidiana, como se observa a continuación,

$$\begin{aligned}
\bar{N}(m'(x), t'(x)) &= \sqrt{\sum_{i=0}^{n-1} N(m_i)^2 + \sum_{i=0}^{n-1} N(t_i)^2} \\
&\leq \sqrt{\sum_{i=0}^{n-1} [N(q) - N(B_m)]^2 + \sum_{i=0}^{n-1} [N(q) - N(B_m)]^2} \\
&\leq \sqrt{\sum_{i=0}^{n-1} [N(q) - N(\text{mín}(B_m, B_t))]^2 + \sum_{i=0}^{n-1} [N(q) - N(\text{mín}(B_m, B_t))]^2} \\
&= \sqrt{2n[N(q) - N(\text{mín}(B_m, B_t))]^2} \\
&= [N(q) - N(\text{mín}(B_m, B_t))]\sqrt{2n} \\
&\leq [\nu(q) - N(\text{mín}(B_m, B_t))]\sqrt{2n},
\end{aligned}$$

es decir, se cumple que,

$$\bar{N}(m'(x), t'(x)) \leq [\nu(q) - N(\text{mín}(B_m, B_t))]\sqrt{2n}, \tag{7.8}$$

pero no todo vector en L^{ET} que satisfaga la condición de norma euclidiana en la desigualdad (7.8) y la condición de congruencia (7.2), será una firma válida. Para simplificar o hacer que sea más fácil para un falsificador potencial realizar una falsificación, supongamos que solo se necesita la condición de norma (7.8), en lugar de la condición de norma infinito (7.3) y (7.4), para obtener una firma falsa.

Además, poniéndonos más a favor del falsificador, supondremos que $B_m = B_t = 0$, de modo que solo se necesita encontrar un vector en $R(q) \times R(q)$. Esto nos da una longitud objetivo τ como se expresa a continuación,

$$\begin{aligned}
\tau = \sqrt{N(f(x))^2 + N(g(x))^2} &\approx \sqrt{N(f(x))^2 + N(f(x))^2}, \quad \text{por que } f(x) \approx g(x) \\
&= \sqrt{2}\sqrt{\sum_{i=0}^{n-1} N(f_i)^2} \\
&\leq \sqrt{2}\sqrt{\sum_{i=0}^{n-1} N(q)^2} \\
&= \sqrt{2n}N(q) \\
&\leq \sqrt{2n}\nu(q).
\end{aligned}$$

Entonces, vemos que τ está dado aproximadamente por la cantidad,

$$\tau = \bar{N}(m'(x), t'(x)), \qquad \text{siempre que } B_m = B_t = 0.$$

Por lo tanto, el defecto gaussiano para nuestro problema τ-aprox-CVP es:

$$\rho = \frac{\tau}{\gamma(L^{ET} \cap\ p\mathbb{Z}[\omega]^{4n})} = \frac{\sqrt{2n}\nu(q)}{\sqrt{\frac{6n}{\pi e}}||q||_2} = \nu(q)\sqrt{\frac{2n\pi e}{6n\nu(q)}} = \nu(q)\sqrt{\frac{\pi e}{3\nu(q)}}, \tag{7.9}$$

que es aproximadamente un múltiplo pequeño de $dim(L^{ET} \cap\ p\mathbb{Z}[\omega]^{4n}) = 4n$, por lo que, la heurística δ-LLL dice que resolver el problema aproximado CVP asociado es un problema computacionalmente difícil de resolver en el lattice $L^{ET} \cap\ p\mathbb{Z}[\omega]^{4n}$.

Una observación importante es que, en teoría, este análisis funciona tanto para polinomios que siguen una distribución gaussiana como para los que siguen una distribución uniforme.

De manera similar como realizó su procedimiento J. Hoffstein en [11], daremos otro cálculo del defecto gaussiano ρ, el cual se basa únicamente en polinomios con distribución uniforme. Tenemos lo siguiente:

El lattice de intersección $L^{ET} \cap\ p\mathbb{Z}[\omega]^{4n}$, es generado por las filas de la matriz,

$$B^{NT} = \begin{bmatrix} \langle p \rangle \langle I_n \rangle & \langle p \rangle \langle H \rangle \\ 0 & \langle pq \rangle \langle I_n \rangle \end{bmatrix},$$

donde p y q son los parámetros seleccionados como en la sección 7.2.1, la matriz $\langle H \rangle$ es la matriz circulante por bloques de la clave pública $h(x)$, $\langle I_n \rangle$ es una matriz identidad por bloques $I_{2\times 2}$. También suponemos que este lattice se comporta como un lattice aleatorio.

En este esquema de firma $ETRUSign$, cada coordenada de los vectores m' y t' se distribuyen de manera aproximadamente aleatoria y uniforme entre la región fundamental H que genera el ideal (q). Ignorando las constantes B_i, es decir, considerando que $B_m = B_t = 0$, el coeficiente cuadrático medio de los coeficientes de (m', t') se calcula de la siguiente manera:

En la siguiente sección se da un cálculo exacto de cuántos números eisenianos z_i hay en H, tales que $Re(z_i) \in H$ y $Im(z_i) = 0$, donde H es una región fundamental del ideal (q), el cual resulta ser aproximadamente igual a $2||q||_2 + 1$.

Consideremos los n coeficientes del polinomio m' o t' como un vector predictor de la forma $(\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_{n-1})$, y sean $(y_0, y_1, \ldots, y_{n-1})$ un vector de los verdaderos valores, cuando se selecciona una muestra de n coeficientes en H para m' o t'. El error cuadrático medio para las partes reales de los $\hat{y}_i$ está dado por,

$$\begin{aligned} ECM_R = \frac{1}{n}\sum_{i=0}^{n-1}(Re(\hat{y}_i) - Re(y_i))^2 &\leq \frac{1}{n}\sum_{i=0}^{n-1}(2||q||_2)^2 \\ &= 4(\sqrt{\nu(q)})^2 \\ &= 4\nu(q). \end{aligned}$$

Análogamente, la ecuación (7.11), de la siguiente sección, muestra un cálculo aproximado para el número de eisenianos z_i, con $Im(z_i)$ dentro de H, el cual resulta ser de aproximadamente

$4r_2/\sqrt{3}$, donde r_2 es el radio en el cual está circunscrito la región H.

Considerando el mismo vector predictor anterior $(\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_{n-1})$, y el vector de los verdaderos valores $(y_0, y_1, \ldots, y_{n-1})$, en una muestra para los coeficientes de m' o t'. Entonces tenemos que el error cuadrático medio para las partes imaginarias de los $\hat{y}_i$ es,

$$\begin{aligned} ECM_I = \frac{1}{n}\sum_{i=0}^{n-1}(Im(\hat{y}_i) - Im(y_i))^2 &\leq \frac{1}{n}\sum_{i=0}^{n-1}\left(\frac{4r_2}{\sqrt{3}}\right)^2 \\ &= \frac{16r_2^2}{3}. \end{aligned}$$

Por lo tanto, las normas euclidianas de los polinomios $m'(x)$ y $t'(x)$ será de aproximadamente,

$$\begin{aligned} \bar{N}(m'(x))^2 \approx \bar{N}(t'(x))^2 \approx \sum_{i=1}^{n} N(m_i)^2 &= \sum_{i=1}^{n}(a_i^2 + b_i^2) \\ &= \sum_{i=1}^{n}\left(4\nu(q) + 16r_2^2/3\right) \\ &= 16n\nu(q) + \frac{16n}{9}\nu(q) \\ &= \frac{52}{9}n\nu(q). \end{aligned}$$

Notemos que la fracción 52/9 es aproximadamente 5,777, por lo que, considerar la cantidad de 6, es una aproximación de la fracción anterior que la acota superiormente, de modo que podemos considerar que,

$$\bar{N}(m'(x))^2 \approx \bar{N}(t'(x))^2 \approx 6n\nu(q).$$

Luego,

$$\rho = \frac{\text{longitud objetivo}}{\text{longitud heuristica Gaussiana}} = \frac{\sqrt{12n\nu(q)}}{\sqrt{6n/\pi e}||q||_2} = \nu(q)\sqrt{\frac{12n}{\frac{6n\nu(q)}{\pi e}}} = \nu(q)\sqrt{\frac{2\pi e}{\nu(q)}}. \quad (7.10)$$

Note que se han propuesto dos posibles valores del defecto gaussiano (7.10) y (7.9), en la práctica es recomendable usar (7.10) ya que los parámetros están mejor ajustados a los parámetros propuestos en la sección 7.2.4.

Por lo tanto, algún atacante puede emplear el ataque anterior y usar técnicas de reducción de lattices para intentar obtener firmas falsas para ciertos mensajes μ_i, con la consideración de que resolver el problema aproximado CVP es computacionalmente difícil.

7.3.4. Probabilidad de fallo de firma

En esta sección veremos un análisis de probabilidad de fallo de firma, es decir, una vez que se han creado o generado posibles firmas válidas por medio de algún método, por ejemplo, mediante el ataque de falsificaciones de la sección 7.3.3; en donde cada firma generada por el atacante tiene una probabilidad de que el algoritmo de verificación del esquema de firma *ETRUSign* produzca como "no válida"para algunas de estás supuestas firmas.

Esto se debe a que si $(m(x), t(x))$ es una posible firma, se debe satisfacer la condición de congruencia (7.2) y las condiciones de norma (7.3) y (7.4) que solicita el algoritmo de verificación. En particular notemos que el ataque de falsificación de firmas de la sección 7.3.3 busca vectores que se encuentran en el lattice de intersección $L^{ET} \cap p\mathbb{Z}[\omega]^{4n}$, así que, cualquier vector que se encuentre en este lattice cumplirá con la condición de congruencia (7.2), pero no siempre la condición de norma (7.3) y (7.4), de modo que podemos estudiar la probabilidad de que los coeficientes de $(m(x), t(x))$ satisfagan la condición de norma requerida.

Por la observación 5.3.1 del capítulo 5, podemos considerar que el vector (m, t) satisface la condición de norma del algoritmo de verificación del esquema $ETRUSign$ si se cumple que cada coeficiente de (m, t) se encuentra dentro de un dominio fundamental generado por el ideal (q), asumiendo que $B_m = B_t = 0$, con $0 = 0 + 0\omega$.

Definición 7.3.1. *Sean x, y variables aleatorias discretas que toman valores $x_1, x_2, \ldots, x_n$ y $y_1, y_2, \ldots, y_m$, respectivamente. El vector aleatorio (x, y) tiene distribución uniforme discreta sobre el conjunto $\{x_1, x_2, \ldots, x_n\} \times \{y_1, y_2, \ldots, y_m\}$ si su función de probabilidad es,*

$$f(x,y) = \begin{cases} 1/nm, & si\ x \in \{x_1, x_2, \ldots, x_n\},\ y \in \{y_1, y_2, \ldots, y_m\} \\ 0, & otro\ \ caso \end{cases}$$

Sea x una variable aleatoria discreta tal que $x = Re(z)$, con $z \in H$, y $Im(z) = 0$, es decir, la variable aleatoria x representa coordenadas de enteros eisenianos que están sobre el eje real y dentro del dominio fundamental H generado por el ideal (q).

Sean y una variable aleatoria discreta tal que $y = Im(z)$, con $z \in H$, es decir, la variable aleatoria y representa coordenadas de enteros eisenianos dentro del dominio fundamental H generado por el ideal (q).

Contamos los elementos del recorrido o rango de la variable aleatoria y. Para cada coordenada real x, de un entero eiseniano $z = c + d\omega$, que se encuentra dentro de la región H, se tiene por lo menos una coordenada imaginaria y_j, de modo que $z = c + d\omega = x + yi$. Estas coordenadas las podemos visualizar en la figura 4.2 del capítulo 4. Llamamos z_1 al vértice que se encuentra más hacia arriba del eje real en sentido positivo del eje imaginario y $-z_1$ al vértice que se encuentra más hacia abajo del eje real en sentido negativo del eje imaginario.

Sean y_s y y_t las coordenadas imaginarias de z_1 y $-z_1$, respectivamente. Por la simetría respecto al origen de la región fundamental H, tenemos que $y_s = y_t$. Luego, consideramos el segmento de la recta $r = 0$ que va desde y_s hasta y_t, vemos que la longitud de la recta r es $2y_s$.

Notemos que en la recta r se encuentran todas las coordenadas imaginarias de los enteros eisenianos z, tal que $z \in H$, dichas coordenadas tienen la forma $c\sqrt{3}/2$, con $c \in \mathbb{Z}$.

Observación 7.3.3. *Por definición de dominio fundamental, H, se tiene que H se encuentra entre los círculos de radio $r_1 = \sqrt{\nu(q)}/2$ y el de radio $r_2 = \sqrt{\nu(q)/3}$. Entonces, se cumple que $y_s \leq r_2$.*

Por lo tanto, es fácil ver que hay un total de,

$$\frac{2y_s}{\sqrt{3}/2} = \frac{4y_s}{\sqrt{3}},$$

coordenadas imaginarias, y_j, en H.

Luego, $|\{y_1, y_2, \ldots, y_m\}| = \dfrac{4y_s}{\sqrt{3}}$ es la cardinalidad del recorrido la variable aleatoria y.

Por la observación 7.3.3, podemos ver que,

$$|\{y_1, y_2, \ldots, y_m\}| = \frac{4y_s}{\sqrt{3}} \leq \frac{4r_2}{\sqrt{3}},$$

es decir, podemos considerar que,

$$|\{y_1, y_2, \ldots, y_m\}| = \frac{4r_2}{\sqrt{3}}, \tag{7.11}$$

es una buena aproximación de la cardinalidad de $|\{y_1, y_2, \ldots, y_m\}|$, de modo que consideraremos a $4r_2/\sqrt{3}$ como la cardinalidad de dicho conjunto.

Ahora, contamos los elementos del recorrido o rango de la variable aleatorio x.

Denotamos por $\mathbb{Z}^+$ al conjunto de los enteros racionales que son positivos y por $\mathbb{Z}^-$ al conjunto de los enteros racionales que son negativos.

Consideremos el subconjunto de números reales $\mathbb{Z} \cup \frac{1}{2}\mathbb{Z}$, donde $\frac{1}{2}\mathbb{Z} = \left\{\frac{t}{2} : t \in \mathbb{Z}\right\}$, podemos definir dos subconjuntos de $\mathbb{Z} \cup \frac{1}{2}\mathbb{Z}$ que serán útiles para los conceptos siguientes,

$$\frac{1}{2}\mathbb{Z}^+ = \left\{\frac{t}{2} : t \in \mathbb{Z}^+\right\} \quad \text{y} \quad \frac{1}{2}\mathbb{Z}^- = \left\{\frac{t}{2} : t \in \mathbb{Z}^-\right\}.$$

Entonces, podemos definir dos relaciones para estos conjuntos como sigue:

Relación 1. Se define la relación para un número real $x \geq \frac{1}{2}$, por el elemento $\lfloor\lfloor x\rfloor\rfloor$ en $\mathbb{Z}^+ \cup \frac{1}{2}\mathbb{Z}^+$ más cercano a x, tal que $\lfloor\lfloor x\rfloor\rfloor \leq x$.

Relación 2. Se define la relación para un número real $x \leq -\frac{1}{2}$, por el elemento $\lfloor\lfloor x\rfloor\rfloor$ en $\mathbb{Z}^- \cup \frac{1}{2}\mathbb{Z}^-$ más cercano a x, tal que $x \leq \lfloor\lfloor x\rfloor\rfloor$.

Como el círculo de radio $r_1 = \sqrt{\nu(q)}/2$ se encuentra inscrito en el hexágono H, entonces r_1 es un punto sobre la recta real que es menor a la coordenada real del vértice de H, que se encuentran más a la derecha del origen en el plano complejo.

Si aplicamos la relación 1 al número real r_1, obtenemos una cota, $\lfloor\lfloor r_1\rfloor\rfloor$, de las coordenadas reales de números eisenianos que se encuentran en el intervalo $(0, \lfloor\lfloor r_1\rfloor\rfloor]$. En esta sección habrá $2\lfloor\lfloor r_1\rfloor\rfloor$ coordenadas que cumplen lo anterior, cuando $\lfloor\lfloor r_1\rfloor\rfloor$ es un entero racional o $2\lfloor\lfloor r_1\rfloor\rfloor - 1$, cuando $\lfloor\lfloor r_1\rfloor\rfloor$ es una fracción.

De manera similar habrá $2\lfloor\lfloor r_1\rfloor\rfloor$ o bien $2\lfloor\lfloor r_1\rfloor\rfloor - 1$ coordenadas que satisfacen lo anterior cuando aplicamos este mismo proceso pero ahora con la relación 2 aplicada al número real $-r_1$.

Note que en el conteo anterior no se ha considerado el caso de la coordenada real 0. Entonces sumando la coordenada 0 al conteo anterior podemos concluir que hay,

$$4\lfloor\lfloor r_1\rfloor\rfloor + 1 \text{ o bien } \; 4\lfloor\lfloor r_1\rfloor\rfloor - 1,$$

coordenadas reales de números eisenianos que se encuentran dentro de H.

Luego, $|\{x_1, x_2, \ldots, x_n\}| = 4\lfloor\lfloor r_1\rfloor\rfloor + 1$ o bien $|\{x_1, x_2, \ldots, x_n\}| = 4\lfloor\lfloor r_1\rfloor\rfloor - 1$.

En la práctica cuando el parámetro $q = a + b\omega \in \mathbb{Z}[\omega]$ es tal que $N(q)$ es grande, las cantidades $4\lfloor\lfloor r_1\rfloor\rfloor + 1$ y $4\lfloor\lfloor r_1\rfloor\rfloor - 1$ son aproximadamente las mismas, es decir, se tiene que $4\lfloor\lfloor r_1\rfloor\rfloor + 1 \approx 4\lfloor\lfloor r_1\rfloor\rfloor - 1$.

Otra observación importante con este mismo supuesto de que $N(q)$ es grande, es que podemos considerar que $\lfloor\lfloor r_1\rfloor\rfloor$ es una buena aproximación de r_1. Por lo tanto, tenemos que la cardinalidad de $|\{x_1, x_2, \ldots, x_n\}|$ será de aproximadamente,

$$\begin{aligned} |\{x_1, x_2, \ldots, x_n\}| &= 4\lfloor\lfloor r_1\rfloor\rfloor + 1 \\ &= 4r_1 + 1 \\ &= 4\sqrt{\nu(q)}/2 + 1 \\ &= 2||q||_2 + 1. \end{aligned}$$

Sustituimos las cardinalidades de los recorridos de las variables aleatorias x e y en la función de probabilidad de la definición 7.3.1, y obtenemos que,

$$f(x,y) = \frac{1}{(2||q||_2 + 1)(4r_1/\sqrt{3})} = \frac{1}{\frac{8\sqrt{\nu(q)}\sqrt{\nu(q)}/2 + 4\sqrt{\nu(q)}/2}{\sqrt{3}}}$$

$$= \frac{\nu(q)\sqrt{3}}{4\nu(q) + 2\sqrt{\nu(q)}}$$

$$= \frac{\sqrt{3\nu(q)}}{4||q||_2 + 2}.$$

Consideremos el siguiente corolario, [[31], capítulo 3, corolario 3.2], el cual establece lo siguiente:

Corolario 7.3.1. *Si A es un evento $A \subset \mathbb{R}^2$,*

$$P(A) = \sum_{(x,y)\in A} f(x,y), \qquad \text{para el caso discreto,}$$

$$P(A) = \iint_A f(x,y)dxdy, \qquad \text{para el caso continuo.}$$

Como la región fundamental, H, de un ideal (q) es un subconjunto que tiene un número finito de números eisenianos, consideramos el caso discreto del corolario anterior para establecer lo siguiente:

Sea A el evento tal que los pares (x, y) están dentro de la región fundamenta H, donde x e y son las mismas variables aleatorias que se definieron anteriormente para la parte real e imaginaria de un número eiseniano en H. Entonces, la probabilidad de que el coeficiente t_j de

$t(x)$ o m_i de $m(x)$ se encuentre en el hexágono H, es:

$$P(A) = P(t_j, m_i \in H) = \sum_{(x,y)\in A} f(x,y) = \sum_{(x,y)\in A} \frac{\sqrt{3\nu(q)}}{4||q||_2 + 2}$$

$$= \frac{\nu(q)\sqrt{3\nu(q)}}{4||q||_2 + 2}$$

$$= \frac{\sqrt{3\nu(q)^3}}{4||q||_2 + 2}.$$

Por lo tanto, si suponemos que los n coeficientes de $t(x)$ y de $m(x)$ son variables aleatorias independientes, la probabilidad de que los coeficientes de cada par de polinomios $(m(x), t(x))$ se encuentren en el hexágono H es de aproximadamente,

$$2P\left(\bigwedge_{j=0}^{n-1} t_j \in H\right) = 2\left(\sum_{(x,y)\in A} f(x,y)\right)^n.$$

Entonces, la probabilidad de que los coeficientes del polinomio $m(x)$ o de $t(x)$ no estén dentro del dominio fundamental H es,

$$1 - 2\left(\sum_{(x,y)\in A} f(x,y)\right)^n = 1 - 2\left(\frac{\sqrt{3\nu(q)^3}}{4||q||_2 + 2}\right)^n. \tag{7.12}$$

Finalmente podemos concluir que si un atacante logra obtener posibles firmas falsas por algún método, por ejemplo, mediante el ataque de falsificaciones de la sección 7.3.3, las cuales serán polinomios $(m(x), t(x))$, entonces estás posibles firmas pueden no cumplir las condiciones de norma (7.3) y (7.4), es decir, que el atacante obtiene polinomios $(m(x), t(x))$ que tienen una probabilidad como en (7.12) de que los coeficientes de $m(x)$ o $t(x)$ no estén en el dominio fundamental H. Esto es equivalente a decir que los polinomios $(m(x), t(x))$ tienen probabilidad (7.12) de no satisfacer las condiciones de norma (7.3) y (7.4).

Observe que a medida que el parámetro n aumenta, la probabilidad anterior, (7.12), aumenta y también cuando $||q||_2$ es grande, la probabilidad ya mencionada también aumenta.

7.4. Seguridad de transcripciones

En esta sección veremos un análisis de seguridad de transcripciones similar al de la sección 6.6.3 del capítulo 6, en el cual consideramos una transcripción de firmas usando el algoritmo de firma digital $ETRUSign$ y se verá que con estás firmas no se logra obtener información adicional que vulnere este esquema de firma. Esto se logrará probando que si un usuario que obtiene firmas válidas mediante el esquema $ETRUSign$, entonces estas firmas se distribuyen uniformemente el lattice de norma acotada $L^{ET}(\nu(q) - N(B_m), \nu(q) - N(B_t))$.

Recordemos que un ataque de transcripción es un método para recuperar la clave privada a partir de una larga lista (transcripción) de firmas de la forma:

$$(firma\ válida_i,\ documento\ hash\ i), \quad i = 1, 2, \ldots \tag{7.13}$$

Daremos una proposición análoga a la proposición 6.6.3 del capítulo 6, la cual afirma que la distribución de firmas se mantiene uniforme sobre el lattice de norma acotada $L^{ET}(\nu(q) -$

$N(B_m), \nu(q) - N(B_t))$. Además, se formula una proposición similar a la ya mencionada. Esta proposición expresa que si un usuario solo conoce la clave pública $h(x)$, solo puede producir una transcripción de pares como en (7.13), que es estadísticamente indistinguible de una transcripción análoga producida utilizando el esquema firma $ETRUSign$, con clave secreta $(f(x), g(x))$.

Iniciamos con el análisis de una transcripción creada usando el esquema de firma $ETRUSign$, con clave secreta $(f(x), g(x))$. La condición de muestreo de rechazo que proporciona este esquema permite probar que las firmas resultantes tengan una distribución uniforme en una cierta región de firmas permitidas.

Establecemos la siguiente notación $A_E = \frac{\sqrt{2}N(p)}{3}[\nu(q) + N(p)]$. Consideramos los pasos 2.2 y 2.3 del algoritmo de firma de $ETRUSign$ para definir una función de firma $(m'(x), t'(x)) = \sigma'(f(x), g(x), m_p(x), t_p(x), r(x))$. Por lo tanto, σ' es una función entre los siguientes conjuntos,

$$\sigma' : \underbrace{pR(1) \times R(1)}_{clave\ privada} \times \underbrace{R(p) \times R(p)}_{documento\ hash} \times \underbrace{R(A_E)}_{r(x)} \longrightarrow \underbrace{L^{ET}(\nu(q) - N(B_m), \nu(q) - N(B_t))}_{firma\ potencial},$$

dada explícitamente por,

$$\sigma'(f(x), g(x), m_p(x), t_p(x), r(x)) = (m_0(x) + a(x)f(x), t_0(x) + a(x)g(x)), \tag{7.14}$$

donde,

$$m_0(x) = m_p(x) + pr(x). \tag{7.15}$$

$$t_0(x) = h(x)m_0(x) \mod q. \tag{7.16}$$

$$a(x) = g^{-1}(x)(t_p(x) - t_0(x)) \mod p. \tag{7.17}$$

Escribiremos,

$$\Omega' = pR(1) \times R(1) \times R(p) \times R(p) \times R(A_E),$$

para representar el dominio de la función σ'. A continuación introduciremos el muestreo de rechazo.
Sea $\Omega = \{f(x), g(x), m_p(x), t_p(x), r(x)\} \subset \Omega'$, tal que:

1. $(m'(x), t'(x)) := \sigma'(f(x), g(x), m_p(x), t_p(x), r(x)) = (m_0(x) + a(x)f(x), t_0(x) + a(x)g(x))$.
2. $||m'(x)||_\infty < \nu(q) - N(B_m), \qquad ||t'(x)||_\infty < \nu(q) - N(B_t)$.
3. $||a(x)f(x)||_\infty \leq N(B_m), \qquad ||a(x)g(x)||_\infty \leq N(B_t)$.

La restricción de σ' a Ω, la denotaremos por σ, por lo que, σ es una función,

$$\sigma : \Omega \to L^{ET}(N(q) - N(B_m),\ N(q) - N(B_t)).$$

Consideraremos dos resultados importantes que a continuación los estudiamos, para después poder demostrar la proposición 7.4.2, la cual dice que toda firma válida para el documento hash $(m_p(x), t_p(x))$ tiene el mismo número de preimagenes en el conjunto $R(A_E)$.

Proposición 7.4.1. *El esquema de firma ETRUSign produce firmas que son verificadas como válidas por el respectivo algoritmo de verificación.*

Demostración. Sea $(m'(x), t'(x), \mu)$ una firma generada por el algoritmo de firma $ETRUSign$ para el mensaje μ.

Verificaremos que $t'(x) \equiv h(x)m'(x) \pmod{q}$. Por definición de $t'(x)$ y $t_0(x)$, tenemos que,

$$t'(x) = t_0(x) + a(x)g(x) = h(x)m_0(x) + a(x)g(x).$$

Como $h(x) = F_q(x)g(x) \mod q$, entonces tenemos que $h(x)f(x) = g(x) \mod q$ y al sustituir a $g(x)$ en las expresiones de $t'(x)$, se obtiene que,

$$\begin{aligned} t'(x) &= t_0(x) + a(x)g(x) = h(x)m_0(x) + a(x)g(x) \\ &\equiv h(x)m_0(x) + h(x)a(x)f(x) \ (\text{m}\acute{o}d\ q) \\ &\equiv h(x)[m_0(x) + a(x)f(x)] \ (\text{m}\acute{o}d\ q) \\ &\equiv h(x)m'(x) \ (\text{m}\acute{o}d\ q). \end{aligned}$$

Ahora, verificaremos que $||m'(x)||_\infty \leq \nu(q) - N(B_m)$ y $||t'(x)||_\infty \leq \nu(q) - N(B_t)$. El paso 3 del algoritmo de firma del esquema $ETRUSign$ asegura que las normas anteriores siempre se satisfacen, ya que es uno de los requisitos para que el algoritmo de firma produzca una firma válida.

Finalmente probaremos que $(m'(x), t'(x)) \equiv (m_p(x), t_p(x)) \ (\text{m}\acute{o}d\ p)$.

Para $m'(x) \equiv m_p(x) \ (\text{m}\acute{o}d\ p)$, tenemos,

$$\begin{aligned} m'(x) &= m_0(x) + a(x)f(x), \qquad \text{por definición de } m'(x) \\ &\equiv m_p(x) + pr(x) + a(x)f(x) \ (\text{m}\acute{o}d\ p), \qquad \text{al sustituir } m_0(x) \\ &\equiv m_p(x) \ (\text{m}\acute{o}d\ p), \qquad \text{al aplicar la reducción } \mod p. \end{aligned}$$

Para $t'(x) \equiv t_p(x) \ (\text{m}\acute{o}d\ p)$, tenemos,

$$\begin{aligned} t'(x) &= t_0(x) + a(x)g(x), \qquad \text{por definición de } t'(x) \\ &\equiv t_0(x) + g^{-1}(x)(t_p(x) - t_0(x))g(x) \ (\text{m}\acute{o}d\ p), \qquad \text{al sustituir } a(x) \\ &\equiv t_0(x) + t_p(x) - t_0(x) \ (\text{m}\acute{o}d\ p), \qquad \text{al operar } g^{-1}(x) \text{ con } g(x). \\ &\equiv t_p(x) \ (\text{m}\acute{o}d\ p). \end{aligned}$$

Por lo tanto, las firmas que produce el esquema de firma $ETRUSign$ se verifican como válidas mediante el algoritmo de verificación correspondiente. □

Denotemos por $\Sigma(f(x), g(x), m'(x), t'(x))$ a la colección:

$$\Sigma(f(x), g(x), m'(x), t'(x)) = \left\{ r(x) \in R(A_E) : \sigma(f(x), g(x), m_{p(x)}, t_p(x), r(x)) = (m'(x), t'(x)) \right\}. \tag{7.18}$$

La clave para contar el tamaño del conjunto $\Sigma(f(x), g(x), m'(x), t'(x))$ es la biyección descrita en el siguiente lema.

Lema 7.4.1. *Sea* $C = \{b(x) \in R(p) : ||b(x)f(x)||_\infty \leq N(B_m) \text{ y } ||b(x)g(x)||_\infty \leq N(B_t)\}$ *y sea* $(m'(x), t'(x)) \in L^{ET}(\nu(q) - N(B_m), \nu(q) - N(B_t))$ *que satisface la condición de congruencia* $(m'(x), t'(x)) \equiv (m_p(x), t_p(x)) \ (\text{m}\acute{o}d\ p)$.

Entonces, la siguiente es una biyección de conjuntos,

$$\begin{aligned} &\Phi : C \to \Sigma(f(x), g(x), m'(x), t'(x)), \\ &b \mapsto \frac{m'(x) - m_p(x)}{p} - b(x)\frac{f(x)}{p}. \end{aligned} \tag{7.19}$$

Demostración. Notemos que:

$$\begin{aligned} m'(x) - m_p(x) &= m_0(x) + a(x)f(x) - m_p(x), \qquad \text{por definición } m'(x) \\ &= m_p(x) + pr(x) + a(x)f(x) - m_p(x), \qquad \text{por definición } m_0(x) \\ &= pr(x) + a(x)f(x) \\ &= pr(x) + a(x)pF(x), \qquad \text{por definición } f(x) \\ &= p[r(x) + a(x)F(x)]. \end{aligned}$$

Vemos que los coeficientes de $m'(x) - m_p(x)$ son múltiplos de p, y de manera similar vemos que $f(x) \in pR(1)$ tiene coeficientes divisibles por p. Además, notemos que el polinomio del lado derecho en (7.19) tiene sus coeficientes en $\mathbb{Z}[\omega]$.

Ahora, necesitamos verificar que $\Phi(b(x)) \in \Sigma(f(x), g(x), m'(x), t'(x))$, lo cual quiere decir que debemos probar que $\Phi(b(x)) \in R(A_E)$ y $\sigma(f(x), g(x), m_p(x), t_p(x), \Phi(b(x))) = (m'(x), t'(x))$.

Realizamos los siguientes cálculos:

En la sección 7.2.1 se ha seleccionado el parámetro $p = 2 + \omega$, como un parámetro fijo, vemos que $\bar{p} = 1 - \omega$ y $\nu(p) = p\bar{p} = 3$. Si hacemos $m'(x) - m_p(x) = s(x)$ y $t(x) = b(x)f(x)$, entonces podemos expresar $s(x) = s_0 + s_1x + \cdots + s_{n-1}x^{n-1}$ y $t(x) = t_0 + t_1x + \cdots + t_{n-1}x^{n-1}$, en donde $s_i = a_i + b_i\omega$ y $t_i = c_i + d_i\omega$ son números eisenianos que son coeficientes de $s(x)$ y $t(x)$, respectivamente. Luego, vemos que,

$$\begin{aligned} \bar{p}s_i &= (1 - \omega)(a_i + b_i\omega) \\ &= (a_i + b_i) + (b_i - a_i + b_i)\omega \\ &= (a_i + b_i) + (2b_i - a_i)\omega. \end{aligned}$$

Análogamente,

$$\begin{aligned} \bar{p}t_i &= (1 - \omega)(c_i + d_i\omega) \\ &= (c_i + d_i) + (d_i - c_i + d_i)\omega \\ &= (c_i + d_i) + (2d_i - c_i)\omega. \end{aligned}$$

Calculamos las normas euclidianas de $\bar{p}s_i$ y $\bar{p}t_i$ elevadas al cuadrado,

$$\begin{aligned} N(\bar{p}s_i)^2 &= [N((a_i + b_i) + (2b_i - a_i)\omega)]^2 = (a_i + b_i)^2 + (2b_i - a_i)^2 \\ &= a_i^2 + 2a_ib_i + b_i^2 + 4b_i^2 - 4a_ib_i + a_i^2 \\ &= 2a_i^2 + 5b_i^2 - 2a_ib_i. \end{aligned}$$

Un cálculo similar al anterior nos permite decir que,

$$N(\bar{p}t_i)^2 = 2c_i^2 + 5d_i^2 - 2c_id_i$$

Por lo tanto, tenemos que,

$$N(\bar{p}s_i)^2 = 2a_i^2 + 5b_i^2 - 2a_ib_i \qquad \text{y} \qquad N(\bar{p}t_i)^2 = 2c_i^2 + 5d_i^2 - 2c_id_i. \tag{7.20}$$

Luego, vemos que,

$$\begin{aligned}
||\Phi(b(x))||_\infty &= \left\| \frac{m'(x) - m_p(x)}{p} \frac{\bar{p}}{\bar{p}} - \frac{b(x)f(x)}{p} \frac{\bar{p}}{\bar{p}} \right\|_\infty = \left\| \frac{\bar{p}[m'(x) - m_p(x)]}{\nu(p)} - \frac{\bar{p}[b(x)f(x)]}{\nu(p)} \right\|_\infty \\
&\leq \frac{1}{\nu(p)} \left(||\bar{p}s(x)||_\infty + ||\bar{p}t(x)||_\infty \right), \qquad \text{por la desigualdad del triángulo} \\
&= \frac{1}{3}[\max_i N(\bar{p}s_i) + \max_i N(\bar{p}t_i)] \\
&= \frac{1}{3}\left[\max_i (2a_i^2 + 5b_i^2 - 2a_ib_i)^{1/2} + \max_i (2c_i^2 + 5d_i^2 - 2c_id_i)^{1/2} \right], \qquad \text{por (7,20)} \\
&\leq \frac{1}{3}\left[\max_i \left((5a_i^2 + 5b_i^2 + 5a_i^2 + 5b_i^5) \right)^{1/2} + \max_i \left((5c_i^2 + 5d_i^2 + 5c_i^2 + 5d_i^2) \right)^{1/2} \right].
\end{aligned}$$

Esta última desigualdad se obtuvo al considerar que un binomio al cuadrado siempre es mayor o igual a cero, y al sumar $3a_i^2 + 4a_i^2 + 4b_i^2$ y $3c_i^2 + 4c_i^2 + 4d_i^2$ en cada sumando respectivo. Luego, como $N(p) = \sqrt{5}$ y por definición de norma euclidiana en los términos $N(\bar{p}s(x))$ y $N(\bar{p}t(x))$, se sigue que,

$$\begin{aligned}
||\Phi(b(x))||_\infty &\leq \frac{1}{3}\left[\max_i \left((5a_i^2 + 5b_i^2 + 5a_i^2 + 5b_i^5) \right)^{1/2} + \max_i \left((5c_i^2 + 5d_i^2 + 5c_i^2 + 5d_i^2) \right)^{1/2} \right] \\
&= \frac{1}{3}[\max_i N(p)(2(a_i^2 + b_i^2))^{1/2} + \max_i N(p)(2(c_i^2 + d_i^2))^{1/2}] \\
&= \frac{N(p)}{3}[\max_i \sqrt{2}N(s_i) + \max_i \sqrt{2}N(t_i)] \\
&= \frac{\sqrt{2}N(p)}{3}[||m'(x) - m_p(x)||_\infty + ||b(x)f(x)||_\infty] \\
&\leq \frac{\sqrt{2}N(p)}{3}[\nu(q) - N(B_m) + N(p) + N(B_m)] \\
&= \frac{\sqrt{2}N(p)}{3}[\nu(q) + N(p)].
\end{aligned}$$

Entonces, $||\Phi(b(x))||_\infty \leq \frac{\sqrt{2}N(p)}{3}[\nu(q) + N(p)] = A_E$. Por lo tanto, $\Phi(b(x)) \in R(A_E)$.

Luego, usando las cuatro fórmulas (7.14-7.17) para calcular la firma σ, entonces se sigue que,

$$\sigma(f(x), g(x), m_p(x), t_p(x), \Phi(b(x))) :$$

$$\begin{aligned}
m_0(x) = m_p(x) + p\Phi(b(x)) &= m_p(x) + p\left(\frac{m'(x) - m_p(x)}{p} - b(x)\frac{f(x)}{p} \right) \\
&= m'(x) - b(x)f(x). \qquad (7.21)
\end{aligned}$$

$$\begin{aligned}
t_0(x) &= h(x)m_0(x) \mod q \\
&\equiv h(x)(m'(x) - b(x)f(x)) \ (\text{m}\acute{o}d\ q) \\
&\equiv h(x)m'(x) - b(x)g(x) \ (\text{m}\acute{o}d\ q), \qquad \text{ya que } h(x) \equiv F_q(x)g(x) \ (\text{mód } q) \\
&\equiv t'(x) - b(x)g(x) \ (\text{m}\acute{o}d\ q), \qquad \text{ya que } \ (m'(x), t'(x)) \in L^{ET}. \qquad (7.22)
\end{aligned}$$

Como $(m'(x), t'(x)) \in L^{ET}(N(q) - N(B_m), N(q) - N(B_t))$, $b(x) \in C$ y por la desigualdad del triángulo tenemos,

$$\begin{aligned} ||m_0(x)||_\infty = ||m'(x) - b(x)f(x)||_\infty &\leq ||m'(x)||_\infty + ||b(x)f(x)||_\infty \\ &\leq \nu(q) - N(B_m) + N(B_m) = \nu(q) \end{aligned}$$

$$\begin{aligned} ||t_0(x)||_\infty = ||t'(x) - b(x)g(x)||_\infty &\leq ||t'(x)||_\infty + ||b(x)g(x)||_\infty \\ &\leq \nu(q) - N(B_t) + N(B_t) = \nu(q), \end{aligned}$$

es decir, la relación (7.22) es una igualdad y no solo una congruencia.

Continuando con el cálculo de $\sigma(f(x), g(x), m_p(x), t_p(x), \Phi(b(x)))$, usamos las relaciones (7.23) y (7.22) para calcular el polinomio $a(x)$,

$$\begin{aligned} a(x) &= g^{-1}(x)(t_p(x) - t_0(x)) \text{ (mód } p) \\ &\equiv g^{-1}(x)t_p(x) - g^{-1}(x)t_0(x) \text{ (mód } p) \\ &\equiv g^{-1}(x)t_p(x) - [g^{-1}t(x) - b(x)g^{-1}(x)g(x)] \text{ (mód } p) \\ &\equiv b(x) \text{ (mód } p). \end{aligned}$$

Notemos que tanto $a(x)$ como $b(x)$ están en $R(p)$, esto nos dice que $a(x) = b(x)$.

Ahora, usamos (7.14) y calculamos lo siguiente,

$$\begin{aligned} \sigma(f(x), g(x), m_p(x), t_p(x), \Phi(b(x))) &= (m_0(x) + a(x)f(x), t_0(x) + a(x)g(x)), \quad \text{por def. de } \sigma \\ &= (m'(x) - b(x)f(x) + a(x)f(x), t'(x) - b(x)g(x) + a(x)g(x)) \\ &= (m'(x), t'(x)), \qquad \text{pues } a(x) = b(x). \end{aligned}$$

En la penúltima igualdad anterior se utilizó la ecuación (7.21) y la relación (7.22), considerada como igualdad. Por lo tanto, directamente de la definición del conjunto $\Sigma(f(x), g(x), m'(x), t'(x))$, (7.18), vemos que,

$$\Phi(b(x)) \in \Sigma(f(x), g(x), m'(x), t'(x)).$$

Por otro lado, tomamos un polinomio $r(x) \in \Sigma(f(x), g(x), m'(x), t'(x))$ y calculamos cuántos polinomios $b(x)$ están en el conjunto C, tales que se satisface que $\Phi(b(x)) = r(x)$. Dado que todos los coeficientes de los polinomios $m'(x) - m_p(x)$ y $f(x)$ son divisibles por p, para facilitar la notación escribiremos,

$$m'(x) - m_p(x) = pS(x) \qquad y \qquad f(x) = pF(x),$$

con $S(x)$ y $F(x)$ polinomios adecuados.

Recordemos que por hipótesis el polinomio $F(x)$ es invertible módulo p. Luego, tenemos que,

$$\begin{aligned} \Phi(b(x)) = r(x) &\iff S(x) - b(x)F(x) = r(x) \\ &\iff b(x) \equiv F^{-1}(x)(S(x) - r(x)) \text{ (mód } p), \end{aligned}$$

con $||b(x)||_\infty \leq N(p)$.

Note que los polinomios $F^{-1}(x), S(x), r(x)$ están determinados de manera única, así que el polinomio dado por $F^{-1}(x)(S(x) - r(x))$ módulo p es único. Por lo tanto, hay exactamente un valor de $b(x) \in C$ polinomios que satisfacen $\Phi(b(x)) = r(x)$, es decir, que el polinomio $b(x)$ es el único elemento del conjunto C que es congruente con $F^{-1}(x)(S(x) - r(x))$ módulo p. Esto prueba que la función Φ es biyectiva, lo que concluye la prueba de este lema. □

Proposición 7.4.2. *La función de firma σ tiene la siguiente propiedad: Para elementos fijos,*

$$\begin{array}{c} \textit{clave privada} \quad (f(x), g(x)) \in pR \times R, \ y \\ \textit{documento hash} \quad (m_p(x), t_p(x)) \in R(p) \times R(p), \end{array}$$

la salida de σ, cuando se consulta sobre el polinomio uniformemente aleatorio $r(x) \in R(A_E)$, se distribuye uniformemente sobre el conjunto:

$$\left\{(m'(x), t'(x)) \in L^{ET}(\nu(q) - N(B_m), \nu(q) - N(B_t)) : (m'(x), t'(x)) \equiv (m_p(x), t_p(x)) \ (\text{m}\acute{o}d\ p)\right\},$$

de firmas válidas para el documento hash $(m_p(x), t_p(x))$.

De manera equivalente el tamaño del conjunto,

$$\left\{r(x) \in R(A_E) : \sigma(f(x), g(x), m_p(x), t_p(x), r(x)) = (m_p(x), t_p(x))\right\},$$

es igual para todos los pares $(m(x), t(x)) \in L^{ET}(\nu(q) - N(B_m), \nu(q) - N(B_t)) : (m'(x), t'(x))$, que satisfacen la condición,

$$(m'(x), t'(x) \equiv (m_p(x), t_p(x)) \ (\text{m}\acute{o}d\ p)).$$

Demostración. Por la proposición 7.4.1, sabemos que $\sigma(f(x), g(x), m_p(x), t_p(x), r(x))$ es congruente con $(m_p(x), t_p(x))$ m*ód* p, es claro que se tiene una probabilidad de cero de generar la firma $(m'(x), t'(x))$ si $(m'(x), t'(x)) \not\equiv (m_p(x), t_p(x))$ (m*ód* p). Así que, asumiremos de ahora en adelante que,

$$(m'(x), t'(x)) \equiv (m_p(x), t_p(x)) \ (\text{m}\acute{o}d\ p), \tag{7.23}$$

siempre se cumple cuando la firma $(m'(x), t'(x))$ es generada con el algoritmo de $ETRUSign$, para algún mensaje μ.

El polinomio $r(x) = r_0 + r_1 x + \cdots + r_{n-1}x^{n-1}$ utilizado para generar una firma se elige de manera aleatoria y uniforme sobre el conjunto $R(A_E)$. Entonces, cualquier coeficiente r_i de $r(x)$ cumple que $N(r_i) \leq A_E$, donde $r_i = a_i + b_i\omega \in \mathbb{Z}[\omega]$, y $N(r_i) = \sqrt{a^2 + b^2}$. Elegir un coeficiente r_i que satisfaga lo anterior es equivalente a elegir un par ordenado (a_i, b_i) de enteros racionales tales que $N(r_i) \leq A_E$. Consideremos el plano complejo determinado por la base $\{1, \omega\}$, al cual llamaremos plano de los enteros de Eisenstein o plano eiseniano, y para fines prácticos llamaremos a los ejes de este plano como x para el eje horizontal e y para el eje vertical. Entonces, determinar cuántos elementos (a_i, b_i) satisfacen lo dicho anteriormente es esencialmente contar los puntos sobre el plano eiseniano que están sobre y dentro del círculo de radio A_E.

Los elementos $r_i = a_i + b_i\omega$ que deseamos contar se muestran en la siguiente figura,

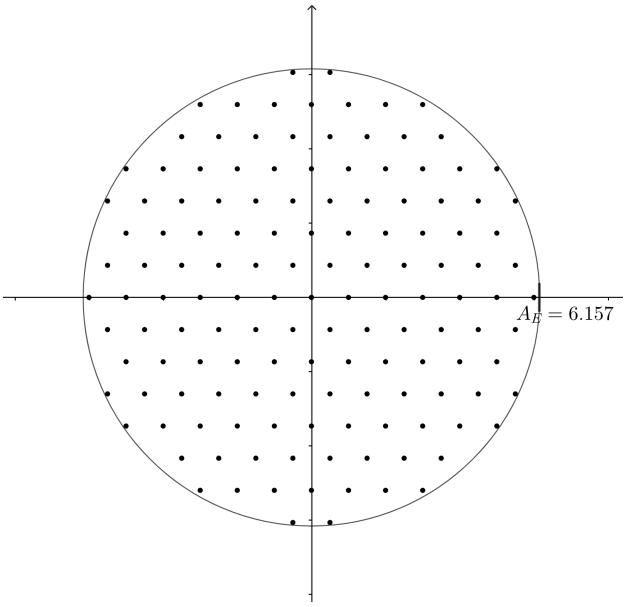


Figura 7.1: Los puntos representan números eisenianos que se encuentran dentro del círculo de radio $A_E = \dfrac{\sqrt{2}N(p)}{3}[\nu(q) + N(p)]$, con valores $p = 2 + \omega$ y $q = 2 + 3\omega$, en el plano eiseniano.

Realizamos el conteo de los elementos $r_i = a_i + b_i\omega$, que están sobre y dentro del círculo de radio A_E de la siguiente manera:

Considere el medio círculo de radio A_E, que se encuentra en la parte positiva del eje y. La ecuación al medio circulo es $C_1 = \sqrt{A_E^2 - x^2}$, mientras que el otro medio círculo que se encuentra en la parte negativa del eje y tiene por ecuación $C_2 = -\sqrt{A_E^2 - x^2}$.

Sea $B = \lfloor A_E \rfloor$. Contamos los r_i que se encuentran en C_1. Claramente en el eje positivo hay B posibles elementos r_i, y de manera similar para el eje negativo x hay B posibles r_i y si incluimos el origen, habrá exactamente $2B + 1$ números eisenianos, $r_i = a_i + b_i\omega$, que están sobre el eje x y que satisfacen estar dentro del círculo $C = C_1 \cup C_2$.

Para el eje y positivo se tienen B posibles r_i que están dentro del círculo C. Ahora bien, para cada $x > 0$ hay $\sqrt{A_E^2 - x^2}$ eisenianos dentro de C. Entonces, en el primer cuadrante incluyendo los ejes coordenados $x > 0$ y $y > 0$, se tiene que hay,

$$(2B + 1) + B + \sum_{i=1}^{B-1} \sqrt{A_E^2 - i^2},$$

elementos r_i. Dado que C es un círculo, tenemos simetría respecto al eje y, y así es fácil ver que en el semiplano positivo hay una cantidad de,

$$(2B + 1) + B + 2\sum_{i=1}^{B-1} \sqrt{A_E^2 - i^2},$$

números eisenianos r_i.

Usando nuevamente el hecho de que el círculo C es simétrico respecto al eje x, tenemos que hay un total de,

$$\begin{aligned}(2B+1)+2\left[B+2\sum_{i=1}^{B-1}\sqrt{A_E^2-i^2}\right] &= 2B+1+2B+4\sum_{i=1}^{B-1}\sqrt{A_E^2-i^2}\\ &= 4B+1+4\sum_{i=1}^{B-1}\sqrt{A_E^2-i^2}\\ &\approx 4A_E+1+4\sum_{i=1}^{B-1}\sqrt{A_E^2-i^2},\end{aligned}$$

eisenianos $r_i = a_i + b_i\omega$ que están dentro del círculo C.

Luego, por el principio de la multiplicación tendremos, aproximadamente,

$$\left(4A_E+1+4\sum_{i=1}^{B-1}\sqrt{A_E^2-i^2}\right)^n,$$

maneras distintas de seleccionar el polinomio $r(x)$.

Por lo tanto, la probabilidad de obtener $(m'(x), t'(x))$ como firma en el documento hash $(m_p(x), t_p(x))$ es igual al número de elementos del conjunto $\Sigma(f(x), g(x), m'(x), t'(x))$ entre $\left(4A_E+1+4\sum_{i=1}^{B-1}\sqrt{A_E^2-i^2}\right)^n$.

Ahora, para todos los documentos hash $(m_p(x), t_p(x))$ que cumplan la condición de congruencia $(m'(x), t'(x)) \equiv (m_p(x), t_p(x))$ (mód p), calculamos la siguiente probabilidad,

$$\begin{aligned}&P_{r(x)\leftarrow R(A_E)}\left(\text{firma es } (m'(x),t'(x)) \mid \text{se tiene clave } (f(x),g(x)) \text{ y hash } (m_p(x),t_p(x))\right)\\ &\quad= \frac{\#\Sigma(f(x),g(x),m'(x),t'(x))}{\#R(A_E)}\\ &\quad= \frac{\#C}{\#R(A_E)},\end{aligned}$$

donde la penúltima igualdad se sigue del lema 7.4.1. Esto completa la demostración de esta proposición. □

Observación 7.4.1. *Notemos que en el esquema de firma NTRUSign la probabilidad de que $(m'(x), t'(x))$ sea firma válida para el documento hash $(m_p(x), t_p(x))$ puede ser mucho mayor que la calculada para el esquema de firma ETRUSign, ya que el espacio total de firmas válidas para NTRUSign es más pequeño que el del esquema ETRUSign para parámetros prácticos, los cuales son de cardinalidad,*

$$(2A+1)^n \qquad \text{y} \qquad \left(4A_E+1+4\sum_{i=1}^{B-1}\sqrt{A_E^2-i^2}\right)^n,$$

respectivamente.

La constante A es menor que la constante A_E, para parámetro prácticos, por lo que, se mantiene la desigualdad $2A < 4A_E$. Además, el factor $4\sum_{i=1}^{B-1}\sqrt{A_E^2-i^2}$ es siempre mayor a cero, y con esto tenemos que,

$$(2A+1)^n < \left(4A_E+1+4\sum_{i=1}^{B-1}\sqrt{A_E^2-i^2}\right)^n.$$

Por lo tanto, cualquier usuario que implemente un ataque de transcripciones contra el esquema $NTRUSign$, que sabemos es poco funcional computacionalmente, y que realice este mismo ataque al esquema $ETRUSign$ también le resultará poco eficiente computacionalmente.

El lema 6.6.1 y el lema 7.4.1 nos permiten ver que tan grandes son los conjuntos C, que se definen en cada lema respectivo, y así, establecer que tan grandes son las probabilidades de tener una firma válida en $NTRUSign$ y $ETRUSign$, respectivamente.

Para dar una versión más completa de la seguridad de transcripciones, necesitamos una versión un poco más fuerte de la proposición 7.4.2, es decir,

Proposición 7.4.3. *La distribución de firmas al consultar σ sobre el documento hash uniformemente aleatorio, de algún mensaje μ, $(m_p(x), t_p(x)) \in R(p) \times R(p)$, es indistinguible de la distribución uniforme sobre el lattice $L^{ET}(\nu(q) - N(B_m), \nu(q) - N(B_t))$.*

La proposición anterior es una consecuencia inmediata de la proposición 7.4.2 bajo el supuesto de que, para cualquier $h(x)$ dada, el número de vectores del lattice de norma acotada, $L^{ET}(\nu(q) - N(B_m), \nu(q) - N(B_t))$, en cada clase lateral $p\mathbb{Z}[\omega]^{4n}$ es esencialmente constante, es decir, que los elementos en el lattice de intersección,

$$L^{ET}(\nu(q) - N(B_m), \nu(q) - N(B_t)) \cap \left[(m_p, t_p) + p\mathbb{Z}[\omega]^{4n}\right],$$

se mantienen distribuidos uniformemente.

Una observación importante es que la proposición anterior no se cumple para algunos lattices. Por ejemplo, con $h(x) = 1$ y observando la matriz,

$$B^{ET} = \begin{bmatrix} \lambda \langle I \rangle & \langle H \rangle \\ 0 & \langle qI \rangle \end{bmatrix},$$

podemos deducir que el lattice L^{ET} tiene rango $4n$, además vemos que por definición la clave pública es $h(x) = F_q(x)g(x) \mod q$ y así,

$$h(x) = 1 = F_q(x)g(x) \mod q,$$

y como $F_q(x) = f^{-1}(x)$, tenemos que $f(x) = g(x) \in pR(1)$. Entonces, para encontrar la clave privada $(f(x), g(x))$, basta con encontrar a $f(x) \in R(1)$, esto se reduce a encontrar un vector en $(m_p, t_p) + p\mathbb{Z}[\omega]^{4n}$, el cual cuenta con solo p^{2n} clases laterales distintas. Es probable que en este esquema se siga cumpliendo la siguiente hipótesis propuesta por J. Hoffstein en [14]:

Suposición 2. Existen constantes reales c, ϵ tal que $\epsilon = 1/poly(n)$ y para todo documento hash $(m_p, t_p) \in R(p) \times R(p)$,

$$\begin{aligned}(1-\epsilon)c &\leq |L^{ET}(\nu(q) - N(B_m), \nu(q) - (B_t)) \cap (m_p, t_p) + p\mathbb{Z}[\omega]^{4n}| \\ &\leq (1+\epsilon)c,\end{aligned}$$

donde $poly(n) = n^{O(1)}$ es la complejidad de tiempo polinomial en n.

Observación 7.4.2. *La suposición 2 nos dice que los elementos en el lattice de intersección $L^{ET}(\nu(q) - N(B_m), \nu(q) - (B_t)) \cap (m_p, t_p) + p\mathbb{Z}[\omega]^{4n}$ mantienen una distribución uniforme para cualquier documento hash $(m_p(x), t_p(x))$.*

Concluimos esta sección señalando que cualquier usuario con acceso a la clave pública, $h(x)$, puede muestrear la distribución uniforme en $L^{ET}(\nu(q)-N(B_m),\nu(q)-N(B_t))$. El usuario simplemente genera aleatoriamente $m(x) \in L^{ET}(\nu(q)-N(B_m))$ hasta que se cumpla la condición,

$$t(x) = h(x)m(x) \in L^{ET}(\nu(q)-N(B_t)).$$

Evidentemente la región de firmas válidas contiene una parte del lattice de norma acotada $L^{ET}(\nu(q)-N(B_m),\nu(q)-N(B_t))$, así que cualquier transcripción generada por cualquier usuario se mantendrá con distribución uniforme en el producto cartesiano $L^{ET}(\nu(q)-N(B_m),\nu(q)-N(B_t)) \times (R(p) \times R(p))$, por la suposición 2.

Por la proposición 7.4.3 y la suposición de que un documento hash es uniforme en $R(p)\times R(p)$, esta transcripción es indistinguible de una transcripción producida por un firmante honesto. La única diferencia entre las dos transcripciones es que el firmante (falsificador) que usó la clave pública $h(x)$, solo conoce los hash $(m_p(x), t_p(x))$ de los mensajes o documentos μ_i.

7.5. Ejemplo de esquema de firma digital $ETRUSign$

Finalizamos este capítulo mostrando un ejemplo del funcionamiento del esquema de firma digital $ETRUSign$, con parámetros pequeños que permiten ver el mecanismo de este esquema de firma digital.

Generación de claves.

Seleccionamos los parámetros:

$$(n, d, k, m) = (7, 1, 3, 4).$$
$$(p, q, B_k, B_m, B_t) = (2+\omega, 2+3\omega, 0+3\omega, 1+\omega, 1+-\omega).$$

Seleccionamos polinomios de manera aleatoria en $T(d+1, d)$:

$$F(x) = x^6 + (\omega^2)x^5 + (\omega)x^4 + x^3 - (\omega)x^2 - \omega^2 x - 1.$$

$$g(x) = x^6 + wx^5 + w^2x^4 + x^3 - wx^2 - wx - 1.$$

Se establece $f(x) = pF(x)$:

$$f(x) = (2+w)x^6 + (2+w)(\omega^2)x^5 + (2+w)(\omega)x^4 + (2+w)x^3 - (2+w)(\omega)x^2 - (2+w)w^2x - (2+w).$$

Los polinomios inversos de $f(x)$ en R_q y $g(x)$ en R_p son,

$$F_q(x) = -(\omega^2)x^6 - (\omega)x^5 - (\omega)x^4 - (\omega^2)x^3 - x^2 - x - \omega.$$
$$g_p(x) = -\omega x^4 - \omega^2 x.$$

Calculamos la clave pública $h(x)$, la cual resulta ser,

$$h(x) = \omega^2 x^6 + \omega^2 x^4 - \omega x^3 + \omega x + \omega^2.$$

Verificamos la condición $NORMF(f(x))$ y $NORMF(g(x))$:

$$F1(x) = \sum_{i=0}^{6} f(x)x^i = (\omega+2)x^6 + (\omega+2)x^5 + (\omega+2)x^4 + (\omega+2)x^3 + \\ + (\omega+2)x^2 + (\omega+2)x + \omega + 2$$

$$= (2+\omega)(x^6 + x^5 + x^4 + x^3 + x^2 + x + 1).$$

$$G1(x) = \sum_{i=0}^{6} g(x)x^i = (-\omega-1)x^6 + (-\omega-1)x^5 + (-\omega-1)x^4 + (-\omega-1)x^3 \\ + (-\omega-1)x^2 + (-\omega-1)x - \omega - 1$$

$$= (-\omega-1)(x^6 + x^5 + x^4 + x^3 + x^2 + x + 1).$$

Entonces, se cumple que:

$$NORMF(F1(x)) = \sqrt{5} \leq \sqrt{9} = N(B_k).$$
$$NORMF(G1(x)) = \sqrt{2} \leq \sqrt{9} = N(B_k).$$

Firmaremos en mensaje: “Este cierre de email formal es una manera segura de terminar muchos diferentes tipos de emails”.

Algoritmo de firma.

Aplicamos la función SHA-3-512 al mensaje a firmar, para así obtener el resumen del mensaje,

$msg\_digest =$ 0011000100110100011000010110010100110100011001100011011001100100
0011010100110011001110000110000100110110001100000110010001100100
0011001001100001011000010011011000110000001100001011000010011100 0
0011100101100100011000100110010100110000001100010011000100110011
0011010100110011011000010011001100110101011000110011011001100010
0011000000110010011001010011011101100110011000100110000101100110
0011100000111001001100010110011000110011001110010011011000110100
0011000101100011001110000110010000110100001101000011000100111001
0011000100110000011001010011000000110010011001000011000101100010
0110000101100011001100110011010001100100011000110110000101100010
0011000000110101011001010110001000111001001110010011100000110101
0110010000110011011000100011010100110111011000100011010000110010
0011000001100100001101110011000101100010011001100110001000110001
0011000000110000001101010011010101100011011001100011000100110011
0011000100110111001100000011010101100001011001000011001101100010
0011010100110111001100010011000001100100011000110011010000110101

El hash del mensaje es,

$$\begin{aligned} m_p(x) &= 3\omega x^3 + 2\omega x^2 + 3\omega x + 2. \\ t_p(x) &= (\omega+2)x^4 + 2\omega x^3 + (\omega+2)x^2 + (2\omega+2)x + 3\omega. \end{aligned}$$

Como los polinomios $m_p(x)$ y $t_p(x)$ están en el anillo R_p, reducimos ambos polinomios módulo p y obtenemos que:

$$\begin{aligned} m_p(x) &= -x^2 - \omega \quad en\ R_p. \\ t_p(x) &= (\omega+1)x^3 + \omega x \quad en\ R_p. \end{aligned}$$

Seleccionamos un polinomio aleatorio en el conjunto $R(A_E)$,

$$r(x) = -\omega x^2 + x + \omega + 1.$$

Calculamos los siguientes polinomios:

$$\begin{aligned} m_0(x) &= 3\omega x^3 + (\omega+1)x^2 + (4\omega+2)x + 2\omega + 3. \\ t_0(x) &= (\omega+1)x^6 + (-\omega-1)x^5 + (\omega+1)x^4 + (-\omega-1)x^2 + (-\omega-1)x + \omega + 1. \\ a(x) &= -x^6 - x^5 + (-\omega-1)x^2 + \omega x. \\ m'(x) &= (\omega+5)x^6 + (-4\omega-2)x^5 + (\omega-4)x^4 + (5\omega+4)x^3 - x^2 + (\omega-1)x + 3\omega - 1. \\ t'(x) &= (2\omega+1)x^6 + (-2\omega-1)x^5 + (\omega-1)x^4 + (\omega+1)x^3 + (2\omega+1)x^2 \\ &\quad + (-2\omega-3)x + 4\omega + 2. \\ a(x)f(x) &= (\omega+5)x^6 + (-4\omega-2)x^5 + (\omega-4)x^4 + (2\omega+4)x^3 + (-\omega-2)x^2 \\ &\quad + (-3\omega-3)x + \omega - 4. \\ a(x)g(x) &= \omega x^6 - \omega x^5 - 2x^4 + (\omega+1)x^3 + (3\omega+2)x^2 + (-\omega-2)x + 3\omega + 1. \end{aligned}$$

Las normas infinito de los siguientes polinomios son,

$$||a(x)f(x)||_\infty = \sqrt{26}, \qquad ||a(x)g(x)||_\infty = \sqrt{13}, \qquad ||m'(x)||_\infty = \sqrt{2}, \qquad ||t'(x)||_\infty = \sqrt{2}.$$

Vemos que las constantes $B_m = 1+5\omega$, $B_t = 3+2\omega$, $\nu(q) - N(B_m) = 7 - \sqrt{26}$ y la constante $\nu(q) - N(B_t) = 7 - \sqrt{13}$, acotan a las cuatro normas anteriores, es decir, se cumple que:

$$\begin{aligned} ||a(x)f(x)||_\infty &= \sqrt{26} \leq \sqrt{26} = N(B_m). \\ ||a(x)g(x)||_\infty &= \sqrt{13} \leq \sqrt{13} = N(B_t). \\ ||m'(x)||_\infty &= \sqrt{2} \leq 7 - \sqrt{26} = \nu(q) - N(B_t). \\ ||t'(x)||_\infty &= \sqrt{2} \leq 7 - \sqrt{13} = \nu(q) - N(B_m). \end{aligned}$$

Por lo tanto, la firma para el mensaje “Existen muchos cifrados resistentes a los ataques de computadoras cuánticas” es:

$$\begin{aligned} m'(x) &= (\omega+5)x^6 + (-4\omega-2)x^5 + (\omega-4)x^4 + (5\omega+4)x^3 - x^2 + (\omega-1)x + 3\omega - 1. \\ t'(x) &= (2\omega+1)x^6 + (-2\omega-1)x^5 + (\omega-1)x^4 + (\omega+1)x^3 + (2\omega+1)x^2 \\ &\quad + (-2\omega-3)x + 4\omega + 2. \end{aligned}$$

Algoritmo de verificación.

Realizamos la primer condición del algoritmo de verificación para el mensaje μ = "Este cierre de email formal es una manera segura de terminar muchos diferentes tipos de emails" y la firma $(m'(x), t'(x))$ calculada en el algoritmo de firma anterior.

Los polinomios $t'(x)$ y $h(x)m'(x)$ son:

$$\begin{aligned} t'(x) &= (2\omega+1)x^6 + (-2\omega-1)x^5 + (\omega-1)x^4 + (\omega+1)x^3 + (2\omega+1)x^2 \\ &\quad + (-2\omega-3)x + 4\omega + 2. \\ h(x)m'(x) &= 10x^6 + (-6\omega-5)x^5 + (8\omega+3)x^4 + (-2\omega+5)x^3 - 7\omega x^2 + x + 7\omega + 7. \end{aligned}$$

La reducción modular de los polinomios $h(x)m'(x)$ y $t'(x)$ en el anillo R_q es:

$$\omega^2 x^6 - \omega^2 x^5 - \omega x^4 - \omega^2 x^3 + \omega^2 x^2 - \omega x.$$

Por lo tanto, se cumple que $t'(x) \equiv h(x)m'(x) \pmod{q}$.

La condición $||m'(x)||_\infty \leq \nu(q) - N(B_m)$ y $||t'(x)||_\infty \leq \nu(q) - N(B_t)$ se cumplen, pues es un requisito para que el algoritmo de firma termine. En este caso se tiene que:

$$\begin{aligned} ||m'(x)||_\infty &= \sqrt{2} \leq 7 - \sqrt{26} = \nu(q) - N(B_t). \\ ||t'(x)||_\infty &= \sqrt{2} \leq 7 - \sqrt{13} = \nu(q) - N(B_m). \end{aligned}$$

Para la condición de congruencia $(m'(x), t'(x)) \equiv (m_p(x), t_p(x)) \pmod{p}$ tenemos que:

$$\begin{aligned} m'(x) &= -x^2 - \omega \quad en\ R_p. \\ t'(x) &= (\omega+1)x^3 + \omega x \quad en\ R_p. \end{aligned}$$

Vemos que $m'(x)$ y $t'(x)$ tienen la misma reducción modular que $m_p(x)$ y $t_p(x)$, respectivamente, y de hecho coinciden ya que el parámetro p solo contiene a las unidades de $\mathbb{Z}[\omega]$ y al elemento 0 como representantes en H_p.

Por lo tanto, se cumple que $(m'(x), t'(x)) \equiv (m_p(x), t_p(x)) \pmod{p}$.

Capítulo 8

Conclusiones

La buena implementación del criptosistema NTRU dada por J. Hoffstein et al. en [14] motivo a seguir con el estudio de un esquema de firma digital basado en NTRU, y fue hasta el año 2018 cuando se logro la estandarización del esquema de firma digital llamado $pqNTRUSign$.

Katherine Jarvis en [20] implemento el criptosistema NTRU sobre el anillo de los enteros de Eisenstein, dando así un nuevo criptosistema llamado ETRU, el cual resulto tener buenas propiedades de seguridad. En esta tesis mostramos parte del trabajo de Katherine Jarvis [20]; principalmente en los capítulo 3, 4, y 5. También, se estudió el esquema de firma digital propuesto por J. Hoffstein et al. en [11]. Ampliamos los análisis y conceptos mencionados anteriormente para diseñar un nuevo esquema de firma digital llamado $ETRUSign$, que tiene un funcionamiento similar al esquema $NTRUSign$ y que brinda una seguridad igual o mayor, cuando tenemos parámetro adecuados, contra los ataques de falsificaciones y los ataques de transcripciones.

Para implementar el esquema de firma $ETRUSign$, estudiamos el esquema de firma digital estándar $pqNTRUSign$ y el esquema $NTRUSign$, junto con un análisis de seguridad para este último. Con la teoría vista en el capítulo 4 y la motivación del criptosistema $ETRU$ se implemento un algoritmo de firma y un algoritmo de verificación, muy similares al de los esquemas descritos anteriormente, sobre el anillo base $\mathbb{Z}[\omega][x]/\langle x^n - 1\rangle$, el cual a simple vista tiene un buen funcionamiento, pero la importancia de saber si al menos el mismo análisis de seguridad que se dio para $NTRUSign$ aplicado al esquema $ETRUSign$ sería adecuado o que cubría las misma afirmaciones que garantiza la seguridad nos llevo a considerar parámetros específicos, por ejemplo $p = 2+\omega$ y $q = a+b\omega$, con $\nu(q)$ un primo racional, etc., y que no limitan la posibilidad cambiar dichos parámetros con restricciones que mantienen amplia posibilidad de elección.

Se revisaron los espacios de claves de los lemas 6.6.1 y 7.4.1 para $NTRUSign$ y $ETRUSign$, respectivamente. En el espacio de claves para $NTRUSign$ J. Hoffstein asegura que la probabilidad de tener una firma válida bajo un ataque de transcripciones es $|C|/(2A+1)^n$, el cual es pequeño, pues $(2A+1)^n$ es amplio; mientras que, en el espacio de claves para $ETRUSign$, obtuvimos una probabilidad de $|C|/(4A_E + 1 + \sum_{i=1}^{B-1}\sqrt{A_E^2 - i^2})^n$. Está última probabilidad también se mantiene pequeña, ya que el término $\left(4A_E + 1 + \sum_{i=1}^{B-1}\sqrt{A_E^2 - i^2}\right)^n$ es mucho más amplio que $(2A+1)^n$, aunque la cardinalidad del conjunto C en $ETRUSign$ es más grande que en $NTRUSign$, con los parámetros de las tablas 7.1 y 6.1, respectivamente.

El objetivo de está tesis se ha logrado, ya que se planteó el esquema de firma $ETRUSign$ junto con un ejemplo para poder firmar algún documento con este esquema, sin embargo, queda por explorar un diseño eficiente del algoritmo de firma y del algoritmo de verificación, así como la búsqueda de nuevos parámetros que mantengan la seguridad de este esquema para tener una

amplia posibilidad de parámetros.

También, el diseño de este esquema, queda como motivación para poder diseñar futuros esquemas de firma digital sobre otros anillos de interés, que utilizan como base lattices similares al lattice NTRU, por ejemplo, sobre los anillos $\mathbb{Z}[\sqrt{-5}]$, $\mathbb{Z}[i]$, en los cuales ya se tiene un criptosistema asociado. Una motivación más es el estudio de las relaciones entre el anillo $\mathbb{Z}[\omega]/(\alpha)$, cuando α es un número de Eisenstein primo, y los campos finitos ya que puede resultar de gran importancia ver qué propiedades de los campos finitos pueden ser aplicadas al anillo $\mathbb{Z}[\omega]/(\alpha)$ para así facilitar la aritmética en el anillo $\mathbb{Z}_q[\omega][x]/\langle x^n - 1\rangle$, dar más seguridad al esquema de firma $ETRUSign$ e incluso brindar nuevas propuestas de implementación para el esquema $NTRUSign$ sobre campos finitos en general.

Apéndice A

Conceptos de Álgebra lineal.

Los conceptos básicos de espacios vectoriales, normas y el algoritmo de Gram-Schmith son fáciles de encontrar en la literatura, por ejemplo, en [33]. Dichos conceptos se presentan en este apéndice, ya que son de gran importancia para entender el problema del vector más corto, el concepto de lattice, y el funcionamiento del algoritmo LLL.

Consideremos espacios vectoriales de dimensión finita, en particular, nos enfocaremos en subespacios del espacio vectorial $\mathbb{R}^m$, con m un entero positivo. Por lo tanto, haremos la convención de que al referirnos a un espacio vectorial, entenderemos que nos referimos al espacio vectorial $\mathbb{R}^m$, a menos que se especifique lo contrario.

Definición A.0.1. *Sea $V \subseteq \mathbb{R}^m$ un subconjunto no vacío de $\mathbb{R}^m$. V es **subespacio vectorial** de $\mathbb{R}^m$ si:*

1. *Para todo $u, v \in V, u + v \in V$.*
2. *Para todo $u \in V,\ \forall \alpha \in \mathbb{R}, \alpha u \in V$.*

Definición A.0.2. *Sean V un espacio vectorial de $\mathbb{R}^n$ y $v_1, v_2, \ldots, v_k$, vectores de V. Una **combinación lineal** de $v_1, v_2, \ldots, v_k$ es un vector $w \in V$ de la forma,*

$$w = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k, \text{ con } \alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{R}.$$

*El conjunto de todas las combinaciones lineales de $v_1, v_2, \ldots, v_k$ es llamado el **espacio generado por** $\{v_1, v_2, \ldots, v_k\}$ y se denota por $gen(v_1, v_2, \ldots, v_k)$.*

Definición A.0.3. *Sea V un espacio vectorial. Los vectores $v_1, v_2, \ldots, v_n \in V$ son **linealmente independientes** si y solo si la única representación del vector 0_V como combinación lineal de los vectores v_i es la trivial, es decir, si:*

$$0_V = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_k, \quad \text{con } \alpha_i \in \mathbb{R},$$

implica que $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$.

Cuando al menos un α_i de la definición anterior es distinto de cero, decimos que los vectores $v_1, v_2, \ldots, v_k \in V$ son **linealmente dependientes**.

Enunciamos la definición de base para un espacio vectorial V junto con algunas de sus propiedades.

Definición A.0.4. *Una **base** para el espacio V es un conjunto de vectores $\{v_1, v_2, \ldots, v_k\}$ contenido en V, tal que $gen(v_1, v_2, \ldots, v_k) = V$ y los vectores $v_1, v_2, \ldots, v_k$ son linealmente independientes.*

Observe que la definición anterior es equivalente a decir que cada vector $w \in V$ se puede escribir de la forma,

$$w = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k,$$

para únicos $\alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{R}$.

Vamos a explicar cómo medir las longitudes en un espacio vectorial, para lo cual se requieren los siguientes conceptos.

Definición A.0.5. *Sean $v, w \in \mathbb{R}^m$, dados por $v = (x_1, x_2, \ldots, x_m)$ y $w = (y_1, y_2, \ldots, y_m)$. Definimos el* ***producto punto****, también llamado* ***producto escalar****, de v con w, por,*

$$v \cdot w = x_1 y_1 + x_2 y_2 + \cdots + x_m y_m.$$

Decimos que dos vectores v y w son *ortogonales* si su producto punto es cero, es decir, si $v \cdot w = 0$.

La *norma Euclidiana* de un vector $v \in \mathbb{R}^m$ está dada por la siguiente relación:

$$||v|| = \sqrt{v \cdot v} = \sqrt{\sum_{i=1}^{m} x_i^2}.$$

Proposición A.0.1. *Sean $v, w \in V \subseteq \mathbb{R}^m$, los cuales suponemos que están anclados al origen 0_v. Si θ es el ángulo que separa a los vectores v y w, entonces:*

1. $|v \cdot w| \leq ||v||\ ||w||$, *Desigualdad de Cauchy Schwarz.*
2. $v \cdot w = ||v||\ ||w||\ cos\theta$.

Demostración. Para la desigualdad de Cauchy Schwarz tenemos que:

Si $v = 0_V$ o $w \neq 0_V$, no hay nada que probar, pues se tiene la igualdad de manera inmediata. Supongamos que ambos vectores v y w son distintos del vector cero. Consideremos la siguiente función,

$$\begin{aligned} f(t) = \| v - tw \|^2 &= (v - tw) \cdot (v - tw) \\ &= v \cdot v - 2tv \cdot w + t^2 w \cdot w \\ &= ||v||^2 - 2tv \cdot w + t^2 ||w||^2. \end{aligned}$$

Vemos que $f(t) \geq 0$, para toda $t \in \mathbb{R}$. Además, el valor de t para el cual se minimiza la función f es, $t = v \cdot w / ||w||^2$. Luego,

$$0 \leq f\left(\frac{v \cdot w}{||w||^2}\right) = ||v||^2 - \frac{(v \cdot w)^2}{||w||^2}.$$

De la desigualdad anterior se sigue que,

$$\frac{(v \cdot w)^2}{||w||^2} \leq ||v||^2.$$

Despejando el término $(v \cdot w)^2$ y tomando valor absoluto de ambos lados, obtenemos la desigualdad de Cauchy Schwarz.

Para probar el segundo resultado de la proposición consideremos lo siguiente:

Ley de los cosenos: Si v, w y $v - w$ son tres vectores que determinan un triángulo arbitrario, se tiene la relación,

$$||v - w||^2 = ||v||^2 + ||w||^2 - 2\ ||v||\ ||w||\ cos\theta.$$

Para visualizar esta ley usando vectores, consideremos el triángulo determinado por los vectores v y w, como se muestra en la siguiente figura.

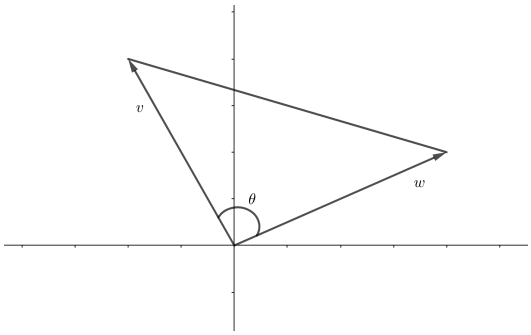


Figura A.1: vectores v y w en $\mathbb{R}^2$.

Entonces, por la ley de los cosenos tenemos que,

$$||v - w||^2 = ||v||^2 + ||w||^2 - 2\ ||v||\ ||w||\ cos\theta, \tag{A.1}$$

ahora, puesto que,

$$||v - w||^2 = (v - w) \cdot (v - w) = ||v||^2 - 2\ v \cdot w + ||w||^2, \tag{A.2}$$

entonces, de las igualdades (A.1) y (A.2), se sigue que,

$$\begin{aligned} ||v||^2 - 2\ v \cdot w + ||w||^2 &= ||u||^2 + ||w||^2 - 2\ ||v||\ ||w||\ cos\theta \\ \implies -2\ v \cdot w &= -2\ ||v||\ ||w||\ cos\theta, \\ o\ bien \qquad v \cdot w &= ||v||\ ||w||\ cos\theta. \end{aligned}$$

□

Definición A.0.6. *Una **base ortogonal de un espacio vectorial** V es una base $\{v_1, v_2, \ldots, v_n\}$, con la propiedad de que,*

$$v_i \cdot v_j = 0, \quad \text{para toda } i \neq j.$$

Además, decimos que la base es ortonormal si $||v_i|| = 1$, para toda i.

Teorema A.0.1. ***(Algoritmo de Gram-Schmidt)**. Sea V un espacio vectorial con producto interno. Si $\{v_1, v_2, \ldots v_n\}$ es una base para V, entonces existe una base de vectores $v_1^*, v_2^*, \ldots, v_n^*$ ortogonales entre sí, por parejas, tales que se cumple la igualdad,*

$$gen(v_1^*, v_2^*, \ldots, v_k^*) = gen(v_1, v_2, \ldots v_k),$$

para toda $k = 1, \ldots, n$. Además, dicha base se construye de la siguiente manera:

Sea $v_1^* = v_1/||v_1||$.

Para $i = 2, \ldots, n$; hacemos $\mu_{ij} = v_i \cdot v_j^*/||v_j^*||^2$, para $1 \leq j < i$ y

$$v_i^* = v_i - \sum_{j=1}^{i-1} \mu_{ij} v_j^*.$$

Demostración. Por inducción sobre k. Para el caso $k = 1$ tenemos que existe un vector $v_1 \in V$, con $v_1 \neq 0_V$. Luego, aplicamos el algoritmo para $i = 1$, de modo que $v_1^* = \dfrac{v_1}{||v_1||}$. Además, es claro que $gen(v_1^*) = gen(v_1)$. Por lo tanto, $\{v_1^*\}$ es una base ortogonal para $gen(v_1)$.

Para el caso $k = 2$ tenemos lo siguiente: si v_1 y v_2 son elementos de V, tales que forman una base para $gen(v_1, v_2)$, y establecemos $v_1^* = v_1$ y $v_2^* = v_2 - \mu_{21} v_1^*$, donde

$$\mu_{2,1} = \frac{v_2 \cdot v_1^*}{||v_1^*||^2}.$$

Entonces, se tiene que,

$$v_1^* \cdot v_2^* = v_1^* \left(v_2 - \frac{v_2 \cdot v_1^*}{||v_1^*||^2} \right) = v_1^* \cdot v_2 - \frac{v_2 ||v_1^*||^2}{||v_1^*||^2} v_1^* = v_1^* \cdot v_2 - v_1^* \cdot v_2 = 0,$$

además, $v_2 = v_2^* + \mu_{21} v_1^*$. Observe que $v_1 = v_1^*$, por lo que, $v_1, v_2 \in gen(v_1^*, v_2^*)$ y por lo tanto, v_1^* y v_2^* son una base de $gen(v_1^*, v_2^*)$, consistente de vectores ortogonales.

Ahora, supongamos que el teorema es válido para $k-1$, es decir, ya que se eligieron vectores ortogonales $v_1^*, \ldots, v_{k-1}^*$, tales que $gen(v_1^*, \ldots, v_{k-1}^*) = gen(v_1, \ldots, v_{k-1})$ y se definieron vectores,

$$v_k^* = v_k - \mu_{k1} v_1^* - \mu_{k2} v_2^* - \cdots - \mu_{k,k-1} v_{k-1}^*,$$

para $\mu_j = v_k \cdot v_j^* / ||v_j^*||^2$. Entonces $v_k^* \cdot v_j^* = 0$, para $j = 1, \ldots, k-1$, ya que

$$\begin{aligned} v_k^* \cdot v_j^* &= v_k \cdot v_j^* - \mu_1 v_1^* \cdot v_j^* - \mu_2 v_2^* \cdot v_j^* - \cdots - \mu_{k-1} v_{k-1}^* \cdot v_j^* \\ &= v_k \cdot v_j^* - \mu_j v_j^* \cdot v_j^*, \quad \text{porque } v_i^* \cdot v_j^* = 0, \quad \text{para } i \neq j \\ &= v_k \cdot v_j^* - \frac{v_k \cdot v_j^*}{||v_j^*||^2} v_j^* \cdot v_j^* \quad \text{por definición de } \mu_j \\ &= 0, \quad \text{ya que } \; v_j^* \cdot v_j^* = ||v_j^*||^2. \end{aligned}$$

Despejando a v_k de la definición de v_k^*, vemos que v_k es combinación lineal de los elementos $v_1^*, \ldots, v_{k-1}^*$, y así se tiene que,

$$gen(v_1, \ldots, v_{k-1}, v_k) = gen(v_1^*, \ldots, v_{k-1}^*, v_k) \subseteq gen(v_1^*, \ldots, v_{k-1}^*, v_k^*).$$

Como v_k^* es combinación lineal, por definición, de v_k y $v_1^*, \ldots, v_{k-1}^*$, se tiene la igualdad en la inclusión anterior. □

Apéndice B

Normas

En este apéndice se estudian las definiciones usuales de norma euclidiana sobre $\mathbb{R}^n$, [22], pero acopladas al anillo de los enteros de Eisenstein y sobre lattices en n dimensiones, que son de utilidad para poder realizar análisis de seguridad para el criptosistema ETRU y los esquemas de firma digital $NTRUSign$ y $ETRUSign$. Se verifica que las propiedades que definen una norma en lattices son las mismas para espacios vectoriales.

Proposición B.0.1. *Sea $\mathbb{Z}[\omega] = \{a + b\omega : a, b \in \mathbb{Z}\}$ el anillo de los enteros de Eisenstein, definimos la función $N : \mathbb{Z}[\omega] \to \mathbb{R}^+ \cup \{0\}$, donde $\mathbb{R}^+$ es el conjunto de los números reales positivos, dada por $N(a + b\omega) = \sqrt{a^2 + b^2}$. Entonces la función N es una norma para el anillo $\mathbb{Z}[\omega]$.*

Demostración. Sean $\alpha = a + b\omega$, $\beta = c + d\omega$ dos números eisenianos cualesquiera y sea $t \in \mathbb{R}$ arbitrario. Probaremos que las siguientes propiedades:

1. $N(\alpha) \geq 0$.

   Vemos que $N(\alpha) = N(a+b\omega) = \sqrt{a^2 + b^2} \geq 0$ ya que $a^2 + b^2 \geq 0$. Por lo tanto $N(\alpha) \geq 0$.

2. $N(\alpha) = 0 \iff \alpha = 0$.

   Supongamos que $N(\alpha) = 0$, es decir, $N(\alpha) = N(a + b\omega) = \sqrt{a^2 + b^2} = 0$. Entonces $a^2 + b^2 = 0$, lo cual ocurre siempre que $a = b = 0$. De este modo se tiene que $\alpha = a + b\omega = 0 + 0\omega$.

   Ahora, supongamos que $\alpha = 0 + 0\omega$. Entonces, $N(\alpha) = N(0 + 0\omega) = \sqrt{0 + 0\omega} = 0$.

3. $N(t\alpha) = |t|N(\alpha)$.

   La igualdad anterior se verifica de la siguiente manera:

$$\begin{aligned} N(t\alpha) = N(t(a + b\omega)) &= \sqrt{(ta)^2 + (tb)^2} = \sqrt{t^2(a^2 + b^2)} \\ &= \sqrt{t^2}\sqrt{a^2 + b^2} \\ &= |t|\sqrt{a^2 + b^2} \\ &= |t|N(\alpha). \end{aligned}$$

4. $N(\alpha + \beta) \leq N(\alpha) + N(\beta)$.

Para esta propiedad probaremos que $N(\alpha+\beta)^2 \leq (N(\alpha)+N(\beta))^2$. Tenemos que,

$$\begin{aligned} N(\alpha+\beta)^2 &= N(a+b\omega+c+d\omega)^2 = N((a+c)+(b+d)\omega)^2 \\ &= \sqrt{(a+c)^2+(b+d)^2} = a^2+2ac+c^2+b^2+2bd+d^2 \\ &= (a^2+b^2)+(c^2+d^2)+2(ac+bd) \\ &= N(\alpha)+N(\beta)+2(ac+bd). \end{aligned}$$

Por otro lado, tenemos que,

$$(N(\alpha)+N(\beta))^2 = (N(\alpha))^2+2N(\alpha)N(\beta)+(N(\beta))^2.$$

Claramente $N(\alpha) \leq N(\alpha)^2$ y $N(\beta) \leq N(\beta)^2$, y probaremos que $2(ac+bd) \leq 2N(\alpha)N(\beta)$.

Vemos que:

$$\begin{aligned} ac+bd \leq N(\alpha)N(\beta) = \sqrt{a^2+b^2}\sqrt{c^2+d^2} &= \sqrt{(a^2+b^2)(c^2+d^2)} \\ &= \sqrt{a^2c^2+a^2d^2+b^2c^2+b^2d^2}. \end{aligned}$$

Sea $x = ac+bd$, $y = \sqrt{a^2c^2+a^2d^2+b^2c^2+b^2d^2}$, probaremos que $x \leq y$.

Caso 1. Supongamos que $x \leq 0$ y $0 \leq y$.
Con estas hipótesis se tiene que $x \leq 0 \leq y$, lo cual implica que $x \leq y$.

Caso 2. Supongamos que $0 < x$ y $0 < y$.
Debemos probar lo siguiente,

$$\begin{aligned} x \leq y &\iff ac+bd \leq \sqrt{a^2c^2+a^2d^2+b^2c^2+b^2d^2} \\ &\iff (ac+bd)^2 \leq (\sqrt{a^2c^2+a^2d^2+b^2c^2+b^2d^2})^2 \\ &\iff a^2c^2+2abcd+b^2d^2 \leq a^2c^2+a^2d^2+b^2c^2+b^2d^2 \\ &\iff 2abcd \leq a^2d^2+b^2c^2. \end{aligned}$$

Luego, considerando la definición de binomio al cuadrado tenemos que,

$$0 \leq (ad-bc)^2 = a^2d^2-2abcd+b^2c^2,$$

lo cual implica que $2abcd \leq a^2d^2+b^2c^2$.

Por lo tanto, $x \leq y$. Entonces, se sigue que $2(ac+bd) \leq 2N(\alpha)N(\beta)$.

Notemos que todos los factores de $N(\alpha+\beta)^2$ son menores o iguales a los de $(N(\alpha)+N(\beta))^2$, es decir, hemos verificado que,

$$N(\alpha+\beta)^2 \leq (N(\alpha)+N(\beta))^2.$$

Finalmente sacando raíz cuadrada de ambos lados de la desigualdad anterior obtenemos que $N(\alpha+\beta) \leq N(\alpha)+N(\beta)$.

Por lo tanto, la función N es una norma para $\mathbb{Z}[\omega]$. □

Proposición B.0.2. *Sea* $\mathbb{Z}[\omega]^n = \{(z_1, z_2, \ldots, z_n) : z_i = a_i + b_i\omega \in \mathbb{Z}[\omega]\}$*, definimos la función* $\bar{N} : \mathbb{Z}[\omega]^n \to \mathbb{R}$*, dada por* $\bar{N}(z_1, z_2, \ldots, z_n) = \sqrt{N(z_1)^2 + N(z_2)^2 + \cdots + N(z_n)^2}$*. Entonces la función* $\bar{N}$ *es una norma para el conjunto* $\mathbb{Z}[\omega]^n$*.*

Demostración. Sabemos que $\mathbb{Z}[\omega]$ y $\mathbb{R}^2$ son isomorfos como $\mathbb{Z}$ módulos. Entonces, podemos dar explícitamente una función que vuelve isomorfos a los a conjuntos anteriores de la siguiente manera:

Definimos la función $\varphi : \mathbb{Z}[\omega] \to \mathbb{R}^2$, dada por $\varphi(a + b\omega) = (a, b)$, entonces se tiene que,

$$N(a + b\omega) = \sqrt{a^2 + b^2} = ||\varphi(a + b\omega)||_2,$$

donde $N(\cdot)$ es la norma euclidiana de $\mathbb{Z}[\omega]$.

Dado que la norma anterior se opera con números enteros, entonces, es fácil ver que la norma usual de $\mathbb{R}^2$ restringida a $\mathbb{Z}^2$, $||\cdot||_2$, sigue siendo norma.

Sean $z_1 = a_1 + b_1\omega, z_2 = a_2 + b_2\omega, \ldots, z_n = a_n + b_n\omega$ números eisenianos, y sean $y_1 = c_1 + d_1\omega, y_2 = c_2 + d_2\omega, \ldots, y_n = c_n + d_n\omega$ números eisenianos.

Definimos la función $\Phi : \mathbb{Z}[\omega]^n \to \mathbb{R}^{2n}$, dada por,

$$\Phi(z_1, z_2, \ldots, z_n) = (\varphi(z_1), \varphi(z_2), \ldots, \varphi(z_n)) = (a_1, b_1, a_2, b_2, \ldots, a_n, b_n)$$

Entonces, tenemos que:

$$\begin{aligned}\bar{N}(z_1, z_2, \ldots, z_n) &= \sqrt{N(z_1)^2 + N(z_2)^2 + \cdots + N(z_n)^2} = \sqrt{a_1^2 + b_1^2 + \cdots + a_n^2 + b_n^2} \\ &= ||\varphi(a_1 + b_1\omega) + \cdots + \varphi(a_n + b_n)||_2,\end{aligned}$$

y como $||\cdot||_2$ es la restricción de la norma usual de $\mathbb{R}^{2n}$ al conjunto $\mathbb{Z}^{2n}$, se tiene que el término $||\varphi(a_1 + b_1\omega) + \cdots + \varphi(a_n + b_n)||_2$ es una norma en $\mathbb{Z}^{2n}$, y de manera similar podemos ver fácilmente que la función Φ es un isomorfismo de $\mathbb{Z}$ módulos. Entonces se sigue $\bar{N}(\cdot)$ es una norma en $\mathbb{Z}[\omega]^n$, también llamada la norma inducida por la norma usual de $\mathbb{Z}^{2n}$.

□

Bibliografía

[1] Abderrahmane, N. , *Cryptanalysis of NTRU with two public keys*, IACR Cryptology ePrint Archive, 2011.

[2] Aggarwal D., Dadush D., Regev O., Stephens-Davidowitz N. *Solving the shortest vector problem in 2n time via discrete Gaussian sampling*. Proceedings of the Annual ACM Symposium on Theory of Computing, pp. 733-742, (2015).

[3] Aguilar Melchor, C., et al. *Sealing the Leak on Classical NTRU Signatures*. PQCrypto 2014, LNCS **8772** Springer (2014), 1-21.

[4] Bandara P., S., *An Exposition of the Eisenstein Integers*. Tesis de Maestría. Eastern Illinois University Charleston, Illinois, 2016.

[5] Backes, W., Wetzel, S. *Parallel Lattice Basis Reduction Using a Multi-threaded Schnorr-Euchner LLL Algorithm.* Euro-Par 2009. Lecture Notes in Computer Science, vol 5704, páginas 960-973 Springer, 2009.

[6] Consortium for Efficient Embedded Securiry, *EESS* #1 : *Implementation Aspects of NTRUEncrypt and pqNTRUSign*, Efficient Embedded Securiry Standars (EESS), 2017, dispobible en: https://csrc.nist.gov/Presentations/2018/NTRUEncrypt-pqNTRUsign

[7] Das, D., Hoffstein, J., Pipher, J. et al. *Modular lattice signatures, revisited*. Des. Codes Cryptogr. 88, 505–532 (2020), https : //doi.org/10,1007/$s10623 - 019 - 00694 - x$

[8] Goldreich, O., Goldwasser, S., Halevi, S. (1997). *Cryptanalysis of the Goldreich-Goldwasser-Halevi Cryptosystem from Crypto '97*. Lecture Notes in Computer Science, vol. 1294, pp. 112-131, Springer-Verlag, 1997.

[9] Griess L. Robert. (2011). *An introduction to groups and lattices: Finite Groups and Positive Definite Rotational Lattices*. International Press.

[10] Hoffstein, J., Pipher, J., Silverman, J. H., *An Introduction to Mathematical Cryptography*, Springer, 2008.

[11] Hoffstein, J., Pipher, J., Schanck, J.M., Silverman, J.H., Whyte, W. *Transcript Secure Signatures Based on Modular Lattices*. In: Mosca, M. (eds) Post-Quantum Cryptography. PQCrypto 2014. Lecture Notes in Computer Science, vol 8772. Springer, Cham. https : //doi.org/10,1007/$978 - 3 - 319 - 11659 - 4_9$

[12] Hoffstein, J., Pipher, J., Silverman, J. H., *NTRU: A Ring-Based Public Key Cryptosystem*. ANTS 1998, vol 1423. Springer, https://doi.org/10.1007/BFb0054868

[13] Hoffstein, J., Pipher, J., Silverman, J.H. NTRU: *A Ring Based Public Key Cryptosystem*. Algorithmic Number Theory (ANTS III), LNCS **1423**, Springer-Verlag (1988), 267-288.

[14] Hoffstein, J., et al. *NTRUSign: Digital Signatures Using the NTRU Lattice*. Topics in Cryptology - CT-RSA 2003. LNCS. **2612**. Springer (2003). 122-140.

[15] Hoffstein, J. and J. Pipher and W. Whyte and Z. Zhang, *A signature scheme from Learning with Truncation*, Cryptology ePrint Archive, 2017.

[16] Howgrave-Graham, N. *A Hybrid Lattice-Reduction and Meet-in-the-Middle Attack Against NTRU*. CRYPTO 2007. Lecture Notes in Computer Science, vol 4622. Springer, Berlin, Heidelberg. (2007) https ://doi.org/10,1007/978 − 3 − 540 − 74143 − 5_9

[17] Ireland, K., and M. Rosen. *A Classical Introduction to Modern Number Theory*, GTM, Springer-Verlag, New York, 1990.

[18] Jaulmes, É., Joux, A. (2000). *A Chosen-Ciphertext Attack against NTRU* CRYPTO 2000. Lecture Notes in Computer Science, vol 1880. Springer, 2000.

[19] Jarvis, K., Nevins, M. *ETRU over the Eisenstein integers*. Des. Codes Cryptogr. **74**, (2015) 219-242.

[20] Katherine Jarvis, NTRU over the Eisenstein Integers. Tesis de Maestría. University of Ottawa. Canadá. 2011.

[21] Lyubashevsky, V., *Lattice Signatures without Trapdoors.*, EUROCRYPT 2012. LNCS, vol 7237 Springer, Berlin, Heidelberg, 2012. https : //doi.org/10,1007/978 − 3 − 642 − 29011 − 4_43

[22] Marsden J., E., Tromba A., J., *Cálculo vectorial*, Addison-Wesley Iberoamericana, 1991.

[23] Micciancio, D., Goldwasser, S., *Complexity of lattice problems a cryptographic perspective*, Springer science+Business Media, LLC, 2002.

[24] Micciancio D., *Introduction to Lattices*, CSE 206A: Lattice Algorithms and Applications, 2010.

[25] Miaoqing, J. *Primality Proving Based on Eisenstein Integers*, Union College, 2016.

[26] Mood A., Graybill F., *Introducción a la teoría de la estadística*, Segunda Edición, ed. Aguilar, 1978.

[27] Murray R. Bremmer, *Lattice Basis Reduction*, An introduction to the LLL Algorithm and Its Applications, first edition, Saskatoon, Canada. A. Chapman and Hall Book (2012).

[28] Nevins, M., KarimianPour, C. and Miri, A. *NTRU over rings beyond* $\mathbb{Z}$. Des. Codes Cryptogr. 56, 65–78 (2010).

[29] Nevins M., KarimianPour C., Miri A., et al *NTRU over rings beyond* $\mathbb{Z}$. Mathematics Subject Classification (2000), Springer Science+Business Media, LLC 2009, 71-73.

[30] Nguyen, Phong Q. *Çripanálisis del criptosistema Goldreich-Goldwasser-Halevi de Crypto '97"* . CRYPTO '99, Londres: Springer-Verlag. págs. 288–304, 1999.

[31] Pérez Salvador, B. R., Castillo Animas, A., De los Cobos Silva, S., *Introducción a la probabilidad*, ed. Universidad Autónoma Metropolitana, 2000.

[32] Pineda Ruelas, M., *Enteros, aritmética modular y grupos finitos*, Primera Edición, ed. Universidad Autónoma Metropolitana, 2014.

[33] Pita Ruiz, Claudio de J., *ÁLGEBRA LINEAL.*, Primera Edición, McGRAW-HILL, 1991.

[34] Rossetti, J. P., *Retículos en espacioa euclídeos*, Universidad Nacional de Córdoba, https://www.famaf.unc.edu.ar/documents/862/BMat48-3.pdf

[35] Wackerly, D., Mendenhall III D., W., L. Scheaffer, R., *Estadística Matemática con aplicaciones*, Cengage Learning, 2010.

[36] Xia, Y., et al *Secure and Efficient Signature Scheme Based on NTRU for Mobile Payment.* J. Phys. Conf. Ser. **910** (2017) 012013, 1-9.

Casa abierta al tiempo
UNIVERSIDAD AUTÓNOMA METROPOLITANA

ACTA DE EXAMEN DE GRADO

No. 00229
Matrícula: 2202800040

ETRUSign: NTRUSign sobre los enteros de Eisenstein

En la Ciudad de México, se presentaron a las 15:00 horas del día 15 del mes de diciembre del año 2022 en la Unidad Iztapalapa de la Universidad Autónoma Metropolitana, los suscritos miembros del jurado:

DR. JUAN CARLOS KU CAUICH
DR. YURIKO PITONES AMARO
DR. JOSE NOE GUTIERREZ HERRERA

Bajo la Presidencia del primero y con carácter de Secretario el último, se reunieron para proceder al Examen de Grado cuya denominación aparece al margen, para la obtención del grado de:

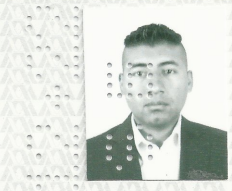
MAESTRO EN CIENCIAS (MATEMÁTICAS APLICADAS E INDUSTRIALES)

DE: OSCAR CASIMIRO MUÑOZ

y de acuerdo con el artículo 78 fracción III del Reglamento de Estudios Superiores de la Universidad Autónoma Metropolitana, los miembros del jurado resolvieron:

Aprobar

Acto continuo, el presidente del jurado comunicó al interesado el resultado de la evaluación y, en caso aprobatorio, le fue tomada la protesta.



OSCAR CASIMIRO MUÑOZ
ALUMNO

REVISÓ

MTRA. ROSALIA SERRANO DE LA PAZ
DIRECTORA DE SISTEMAS ESCOLARES

DIRECTOR DE LA DIVISIÓN DE CBI

Román Linares Romero
DR. ROMAN LINARES ROMERO

PRESIDENTE

DR. JUAN CARLOS KU CAUICH

VOCAL

DR. YURIKO PITONES AMARO

SECRETARIO

DR. JOSE NOE GUTIERREZ HERRERA