MATHEMATICS OF COMPUTATION
Volume 87, Number 310, March 2018, Pages 821–854
http://dx.doi.org/10.1090/mcom/3227
Article electronically published on August 15, 2017

NON-MINIMALITY OF THE WIDTH-w NON-ADJACENT FORM IN CONJUNCTION WITH TRACE ONE τ-ADIC DIGIT EXPANSIONS AND KOBLITZ CURVES IN CHARACTERISTIC TWO

DANIEL KRENN AND VOLKER ZIEGLER

ABSTRACT. This article deals with redundant digit expansions with an imaginary quadratic algebraic integer with trace ± 1 as base and a minimal norm representatives digit set. For $w \geq 2$ it is shown that the width-w non-adjacent form is not an optimal expansion, meaning that it does not minimize the (Hamming) weight among all possible expansions with the same digit set. One main part of the proof uses tools from Diophantine analysis, namely the theory of linear forms in logarithms and the Baker–Davenport reduction method.

Part I. The beginning

1. INTRODUCTION

Let τ be an (imaginary quadratic) algebraic integer and $\mathcal{D}$ a finite subset of $\mathbb{Z}[\tau]$ including zero. Choosing the digit set $\mathcal{D}$ properly, we can represent $z \in \mathbb{Z}[\tau]$ by a finite sum

$$\sum_{\ell=0}^{L-1} \sigma_\ell \tau^\ell,$$

where the digits σ_ℓ lie in $\mathcal{D}$. We call this representation a digit expansion of z. Using a redundant digit set $\mathcal{D}$, i.e., taking more digits than needed to represent all elements of $\mathbb{Z}[\tau]$, each element can be written in different ways. Of particular interest are expansions which have the lowest number of non-zero digits. We call those expansions *optimal* or *minimal expansions*.

The motivation looking at such expansions comes from elliptic curve cryptography. There the scalar multiplication of a point on the curve is a crucial operation and has to be performed as efficiently as possible. The standard double-and-add algorithm can be extended by windowing methods; see for example [6, 8, 18, 25, 26]. Translating this into the language of digit expansions means the usage of redundant digit expansions with base 2. However, using special elliptic curves, for example Koblitz curves (see [14, 15, 25, 26]), the "expensive" doublings can be replaced by

Received by the editor April 6, 2016 and, in revised form, October 1, 2016.

2010 *Mathematics Subject Classification.* Primary 11A63, 11Y50, 11D75.

Key words and phrases. τ-adic expansions, redundant digit sets, elliptic curve cryptography, Koblitz curves, Frobenius endomorphism, scalar multiplication, Hamming weight, linear forms in logarithms, geometry of numbers, Baker–Davenport method, continued fractions.

The first author was supported by the Austrian Science Fund (FWF): I1136, by the Austrian Science Fund (FWF): P24644-N26, and by the Austrian Science Fund (FWF): W1230, Doctoral Program "Discrete Mathematics".

The second author was supported by the Austrian Science Fund (FWF): P24801.

the "cheap" application of the Frobenius endomorphism over finite fields. In the world of digit expansions this means taking an imaginary quadratic algebraic integer as base. This leaves us with the additions of points of the elliptic curve as an "expensive" operation. The number of such additions is basically the number of non-zero digits in an expansion. Therefore minimizing this number is an important goal.

We are now going back to expansions with a low number of non-zero digits. Let the parameter $w \geq 2$ be an integer. Then one special expansion is the width-w non-adjacent form, where in each block of width w at most one digit is not equal to zero; see Reitwiesner [22] who introduced this notion for $w = 2$ and others including Muir and Stinson [19] and Solinas [25, 26]. It will be abbreviated by w-NAF. This expansion contains, by construction, only few non-zero digits. When we use a digit set consisting of zero and of representatives with minimal norm of the residue classes modulo τ^w excluding those which are divisible by τ, then the w-NAF-expansion is optimal (minimal) in a lot of cases. For example, using an integer (absolute value at least 2) as base τ, the w-NAF is a minimal/optimal expansion; see Reitwiesner [22], Jedwab and Mitchell [13], Gordon [8], Avanzi [1], Muir and Stinson [19], and Phillips and Burgess [21]. As a digit set it contains in these cases zero and all integers with absolute value smaller than $\frac{1}{2}\left|\tau\right|^w$ and not divisible by τ.

A general criterion for optimality of the w-NAF-expanions can be found in Heuberger and Krenn [12]: The w-NAF of each element is optimal, if expansions of weight 2 are optimal. This is especially useful if the digit set has some underlying geometric properties as it is the case for a minimal norm representatives digit set. In Heuberger and Krenn [11] an optimality result for a general algebraic integer base is given. A refinement of this general criterion in the imaginary quadratic case is stated in [12]. For τ being imaginary quadratic and a zero of $\tau^2 - p\tau + q$, the main result is that optimality follows if $|p| \geq 3$ and $w \geq 4$. Further, there are conditions given for $w = 2$ and $w = 3$. In the cases $p = \pm 2$ and $q = 2$ the w-NAF-expansion is optimal for odd w and non-optimal for even w. Moreover, non-optimality was also shown when $p = 0$ and w is odd.

In this article we are interested in the case when $p \in \{-1, 1\}$. Note that the case $q = 2$ is related to Koblitz curves in characteristic 2; see Koblitz [14], Meier and Staffelbach [17], and Solinas [25, 26]. A few results are already known: If $w = 2$ or $w = 3$ optimality was shown in Avanzi, Heuberger and Prodinger [2, 3] (see also Gordon [8] for $w = 2$). In contrast, for $w \in \{4, 5, 6\}$, the w-NAF-expansion is not optimal anymore. This was shown in Heuberger [9]. These results rely on transducer automata rewriting arbitrary expansions (with given base and digit set) to a w-NAF-expansion and on a search of cycles of negative "weight".

Experimental results checking the above criterion by symbolic calculations indicate that the w-NAF is non-optimal for $w \geq 4$ and, moreover, non-optimal for all $w \geq 2$ when $q \geq 3$; see Heuberger and Krenn [12]. The main result presented in this work—see the next section for a precise statement—proves this conjecture for $q \leq 500$.

2. Expansions, the results and an overview

We use this section to present our main theorem and to give an overview of the different methods used during its proof. We start by explaining what we mean by optimal (or minimal) expansions.

Definition 2.1. Let τ be an algebraic integer, and suppose we have a set $\mathcal{D}$ (called *digit set*) with $\mathcal{D} \subseteq \mathbb{Z}[\tau]$ such that $0 \in \mathcal{D}$. Let w be an integer with $w \geq 2$ (called *window size*).

(1) The finite sum
$$z = \sum_{\ell=0}^{L-1} \sigma_\ell \tau^\ell$$
with a positive integer L and $\sigma_\ell \in \mathcal{D}$ for all ℓ is called a *digit expansion of* z with base τ.

(2) We call the expansion defined above a *width-w non-adjacent form* (abbreviated by *w-NAF*) if for $\ell \in \{0, 1, \ldots, L-w\}$ each of the words
$$\sigma_\ell \sigma_{\ell+1} \ldots \sigma_{\ell+w-1}$$
contains at most one non-zero digit.

(3) The number of non-zero digits is called the *(Hamming) weight* of the expansion.

(4) Suppose we have an expansion of z with weight W. We call this expansion *optimal* or *minimal* if each expansion of z (with digits out of $\mathcal{D}$) has a weight which is at least W.

(5) The w-NAF expansion is said to be *optimal* or *minimal* (with respect to τ and to $\mathcal{D}$) if the w-NAF of each element of $\mathbb{Z}[\tau]$ is minimal.

We will skip "with respect to τ and to $\mathcal{D}$" in the previous definitions if this is clear from the context (and in our cases it will always be the base τ and the minimal norm digit set $\mathcal{D}$).

Before we are able to state our main theorems, we have to specify the digit set $\mathcal{D}$. For a parameter w (the window size) we assume that 0 is a digit and that we take a representative of minimal norm out of each residue class modulo τ^w which is not divisible by the base τ. We call such a digit set a *minimal norm representative digit set modulo τ^w*; see section 12, in particular Definition 12.1, for details.

Remark 2.2. If τ is an imaginary quadratic algebraic integer (as we use it here in this article) and $\mathcal{D}$ a minimal norm representative digit set modulo τ^w, then each element of $\mathbb{Z}[\tau]$ admits a unique w-NAF expansion; see Heuberger and Krenn [10].

Now it is time to state our main results.

Theorem 2.3. *Let q be an integer with $q \geq 2$ and let $p \in \{-1, 1\}$. Let τ be a root of $X^2 - pX + q$ and $\mathcal{D}$ a minimal norm representative digit set modulo τ^w. Then there exists an effectively computable bound w_q such that for all $w \geq w_q$ the width-w non-adjacent form expansion is not optimal with respect to τ and to $\mathcal{D}$. In particular, we may choose*[1]
$$w_q = 8.68 \cdot 10^{15} \log q \log\log q \quad \textit{if } q \geq 13$$
and
$$w_q = 1.973 \cdot 10^{16} \quad \textit{if } q \in \{2, 3, \ldots, 12\}.$$

It turns out that the bounds are rather huge (section 7). However, for small q we can reduce this bound dramatically (for example from $w_2 = 8.596 \cdot 10^{15}$ to $\widetilde{w}_2 = 140$) and get the following much stronger result.

[1]The explicit bounds w_q (Theorem 2.3, "in particular"-part) are rough estimates. For a particular q, better bounds can be computed, which is done throughout this article.

Theorem 2.4. *Let q and w be integers with*

- *either*[2] $2 \leq q \leq 500$ *and* $w \geq 2$
- *or* $q \geq 2$ *and* $w \in \{2, 3\}$,

and let $p \in \{-1, 1\}$. Let τ be a root of $X^2 - pX + q$ and $\mathcal{D}$ a minimal norm representative digit set modulo τ^w. Then the width-w non-adjacent form expansion with respect to τ and to $\mathcal{D}$ is optimal if and only if $q = 2$ and $w \in \{2, 3\}$.

Formulated differently, this means that the w-NAF is **not** minimal/optimal for all the given parameter configurations except for the four cases with $w \in \{2, 3\}$, $p \in \{-1, 1\}$ and $q = 2$.

The main part of the proof of Theorem 2.4 deals with an algorithm which takes q (and p) as input and outputs a list of values for w for which no counterexample to the minimality of the w-NAF was found. Let us also formulate this as a proposition.

Proposition 2.5. *Let q be an integer with $q \geq 2$ and $p \in \{-1, 1\}$. Let τ be a root of $X^2 - pX + q$ and $\mathcal{D}$ a minimal norm representative digit set modulo τ^w. Then there is an algorithm which tests non-optimality of the width-w non-adjacent form expansion for all $w \geq 2$.*

This algorithm grows out of an intuition on how a counterexample to minimality of the w-NAFs is constructed. To do so, we have to find certain lattice point configurations located near the boundary of the digit set. This is described in general in the overview (section 11) of Part III and with more details and very specific for our situation in section 16. This leaves us to find a lattice point located in some rectangle which additionally avoids some smaller lattice.

All of Part II deals with this problem of finding a suitable lattice point inside the given rectangle, which is precisely formulated as Proposition 5.1. Using the theory of the geometry of numbers allows us to construct such a lattice point, but unfortunately not "for free"; we have to ensure that there is no lattice point in some smaller rectangle. This problem can be reformulated as an inequality (namely inequality (5.1)) and we have to show that it does not have any integer solutions.

Dealing with the solutions of inequality (5.1) is a task of Diophantine analysis. In particular and because of the structure of (5.1a) we use the theory of linear forms in logarithms. This provides, for given q, a rather huge bound on w (due to Matveev [16]); see section 7 for details. From this Theorem 2.3 can be proven. However, using the convergents of continued fractions we are able to reduce this bound significantly. Therefore, we are able to check all the remaining w directly. In particular, we use a variant of the Baker–Davenport method [5], which is described in section 8.

Let us close this overview of the second part with the following: In section 6 we give some remarks on how to test inequality (5.1) directly. The actual algorithm is stated in section 10.

In Part III digits come into play and the counterexamples to minimality of the w-NAF are constructed. Section 13 explains this directly for some values of w (but arbitrary q), whereas the remaining sections deal with the construction using the result of Part II. In particular, this gives us a minimal non-w-NAF expansion, whose most significant digit is perturbated a little bit (section 14). This is compensated by

[2]The upper bound $q \leq 500$ in Theorem 2.4 is determined by an extensive computation. Details can be found at `http://www.danielkrenn.at/koblitz2-non-optimal`, in particular file `result_overview`.

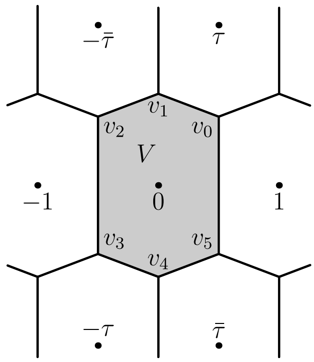

FIGURE 3.1. Voronoi cell V of 0 corresponding to the set $\mathbb{Z}[\tau]$ with $\tau = \frac{1}{2} + \frac{i}{2}\sqrt{7}$ (i.e., $p = 1$, $q = 2$).

a large change in the least significant digit; see section 15. In sections 16 and 17 all results are glued together and the actual counterexamples are constructed (thereby proving Theorem 2.4). The actual algorithm is implemented[3] in SageMath [24].

We are now finished with the introductory overview and will start with two preparatory sections.

3. The set-up

This section is to state a couple of definitions used throughout this work and to fix some notation.

- Let q be an integer with $q \geq 2$. We call q the *norm of our base* (for what we mean by "base", have a look below).
- Let $p \in \{-1, 1\}$. We call p the *trace of our base.*
- Let τ be a zero of $X^2 - pX + q$, more precisely, we take
$$\tau = \frac{p}{2} + i\sqrt{q - \frac{1}{4}}.$$
We call τ the *base of our expansions.*

  Note that the case $\tau = \frac{1}{2}p - i\sqrt{q - \frac{1}{4}}$ (i.e., taking the negative square root) is, by conjugation, "equivalent" to our set-up. This shall mean by constructing a counterexample to minimality of the w-NAF in one case (sign of the square root) and by conjugating everything, we obtain a counterexample for the other sign of the square root.
- Our expansions live in the lattice
$$\mathbb{Z}[\tau] = \{a + b\tau : a \in \mathbb{Z},\, b \in \mathbb{Z}\}.$$
See also section 4 for the other lattices used.
- We set
$$V = \{z \in \mathbb{C} : \forall y \in \mathbb{Z}[\tau]\colon\ |z| \leq |z - y|\}$$
and call it the *Voronoi cell of* 0 corresponding to the set $\mathbb{Z}[\tau]$. An example of this Voronoi cell in a lattice $\mathbb{Z}[\tau]$ is shown in Figure 3.1.

[3]See `http://www.danielkrenn.at/koblitz2-non-optimal` for the code.

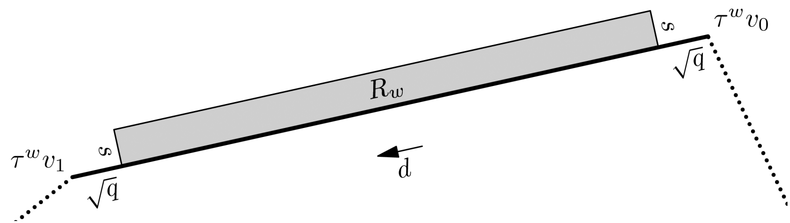

FIGURE 3.2. Rectangle R_5 for $q = 5$ and $p = 1$.

- The *vertices of* V are

$$v_0 = \frac{p}{2} + \frac{i}{2\,\mathrm{Im}(\tau)}\left(\mathrm{Im}(\tau)^2 - \frac{1}{4}\right) = \frac{p}{2} + \frac{i}{\sqrt{4q-1}}\left(q - \frac{1}{2}\right),$$
$$v_1 = \frac{i}{2\,\mathrm{Im}(\tau)}\left(\mathrm{Im}(\tau)^2 + \frac{1}{4}\right) = \frac{i}{\sqrt{4q-1}}q,$$
$$v_2 = -\frac{p}{2} + \frac{i}{2\,\mathrm{Im}(\tau)}\left(\mathrm{Im}(\tau)^2 - \frac{1}{4}\right) = -\frac{p}{2} + \frac{i}{\sqrt{4q-1}}\left(q - \frac{1}{2}\right),$$

  and $v_3 = -v_0$, $v_4 = -v_1$ and $v_5 = -v_2$; see Heuberger and Krenn [10].
- Let w be an integer with $w \geq 2$. We call w the *window size* of our expansions; see also Definition 2.1.
- Let

$$d = \frac{v_1 - v_0}{|v_1 - v_0|}$$

  be the direction from v_0 to v_1, and note that we have

$$d = pi\frac{\tau}{\sqrt{q}}.$$

  Set $d_w = d(\tau/\sqrt{q})^w$. See also Figure 3.2.
- Let

$$s = \sqrt{\frac{q - 1/4}{q+2}}$$

  be the *height of the rectangle* defined below. See also Figure 3.2.
- Define the open *rectangle* R_w with vertices
  - $\tau^w v_0 + d_w\sqrt{q}$,
  - $\tau^w v_1 - d_w\sqrt{q}$,
  - $\tau^w v_1 - d_w\sqrt{q} - pid_w s$, and
  - $\tau^w v_0 + d_w\sqrt{q} - pid_w s$.

  An example of this rectangle is shown in Figure 3.2.

  Note that one side length of the rectangle R_w is

$$|\tau^w v_1 - \tau^w v_0| - 2\sqrt{q} = \sqrt{q}^w \frac{1}{2}\sqrt{1 + \frac{1}{4q-1}} - 2\sqrt{q} = \sqrt{\frac{q^{w+1}}{4q-1}} - 2\sqrt{q} \tag{3.1}$$

  and the other is s.

We finish this section with a couple of remarks.

Remark 3.1. The location of the rectangle R_w (in relation to the scaled Voronoi cell $\tau^w V$) is in such a way that $\tau^{-1}R_w$ has empty intersection with $\tau^w V$. When constructing the actual counterexample in Part III, this will guarantee us that an element of R_w does not become a digit (during division by τ).

Remark 3.2. Note that the rectangle R_w is well defined (has positive area) if and only if $w > \log(16q-4)/\log q$, which follows from positivity of (3.1). Therefore, for $q=2$ we have $w \geq 5$, for $q=3$ we have $w \geq 4$, for $4 \leq q \leq 15$ we have $w \geq 3$, and for $q \geq 16$ we have $w \geq 2$.

4. Lattices

As mentioned above, our digit expansions live in the lattice

$$\mathbb{Z}[\tau] = \{a + b\tau : a \in \mathbb{Z},\, b \in \mathbb{Z}\} = \langle 1, \tau \rangle.$$

It will become handy to define a few other (related) lattices. Our first one is

$$\Lambda_\tau = \langle \tau, q - p\tau \rangle = \langle \tau, \tau^2 \rangle,$$

where we interpret the complex plane embedded into $\mathbb{R}^2$ in the usual way. This lattice is used during the construction of our counterexamples, since we need points divisible by τ there. Further, we also work with the smaller lattice

$$\Lambda_{\tau^2} = \langle \tau^2, \tau^3 \rangle = \langle q\tau, \tau^2 \rangle \subseteq \Lambda_\tau,$$

since in view of Proposition 5.1 we want to avoid this lattice.

Moreover, let us note that the middle point of the lower long side of the rectangle R_w is

$$\frac{v_0 + v_1}{2}\tau^w = \frac{\tau^{w+1}}{2}.$$

In general this is not a point of the lattice Λ_τ but of the lattice

$$\Lambda_{\tau/2} = \frac{1}{2}\Lambda_\tau = \langle \frac{\tau}{2}, \frac{\tau^2}{2} \rangle \supseteq \Lambda_\tau.$$

This is the reason why we will work mainly in the larger lattice $\Lambda_{\tau/2}$.

We need some basic properties of the lattices above. Let us start with Λ_τ and some divisibility conditions for its elements.

Lemma 4.1. *The elements of $\mathbb{Z}[\tau]$ divisible by τ once (i.e., divisible by τ but not by τ^2) are exactly the elements*

$$a\tau + b(q - p\tau)$$

with $a \in \mathbb{Z}$ but $q \nmid a$ and with $b \in \mathbb{Z}$.

Moreover, if $z \in \mathbb{Z}[\tau]$ is a lattice point divisible by τ, then $z-1$, $z+1$, $z+\tau-p$ and $z-\tau+p$ are not divisible by τ.

The structures described above can be found in Figure 4.1.

Proof of Lemma 4.1. An element $c + d\tau \in \mathbb{Z}[\tau]$ is divisible by τ if and only if $q \mid c$. Therefore, if $z \in \mathbb{Z}[\tau]$ is divisible by τ, then $z-1$, $z+1$, $z+\tau-p$ and $z-\tau+p$ are not divisible by τ, which proves the second part of the lemma. As $\tau \nmid a - b\tau$ if and only if $q \nmid a$, multiplying by τ yields the first part of the lemma. $\square$

For analyzing the lattice $\Lambda_{\tau/2}$ and the scaled Voronoi cell $\tau^w V$, it is important to know some arithmetic properties of τ. We show the following lemma to get some insights.

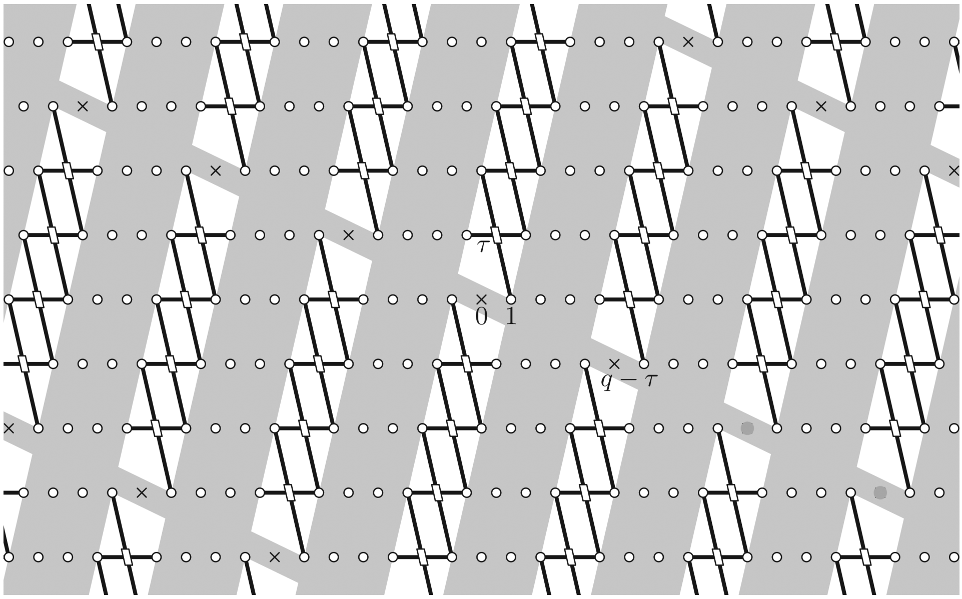

FIGURE 4.1. Lattice $\mathbb{Z}[\tau]$. Points marked with a circle are not divisible by τ, points marked with a rectangle are divisible exactly once by τ, points marked with a cross are divisible by τ^2. There are lines from each rectangle (divisible exactly once by τ) to its neighboring circles (not divisible by τ).

Lemma 4.2. *The algebraic integer τ satisfies the following properties.*

(1) *Every prime $\ell \in \mathbb{Z}$ with $\ell \mid q$ splits in $\mathbb{Z}[\tau]$.*
(2) *If a and b are integers with $a + b\tau = \tau^w$ or $a + b\tau = \bar{\tau}^w$, then* $\gcd(a, b) = 1$.

Note that $\mathbb{Z}[\tau]$ is the maximal order of $\mathbb{Q}(\tau)$.

Proof of Lemma 4.2. The first statement is a direct consequence from algebraic number theory, in particular the splitting of a prime ℓ in $\mathbb{Z}[\tau]$ is described by the factorization of the minimal polynomial of τ mod ℓ (for example, see Theorem 2 in Chapter 11 of Ribenboim [23]). Indeed we have

$$X^2 - pX + q \equiv X^2 \pm X \equiv X(X \pm 1) \mod \ell,$$

hence all primes $\ell \mid q$ split completely in $\mathbb{Z}[\tau]$.

The second statement is trivial for $w \in \{0, 1\}$ (note that we use $w \geq 2$ throughout this paper anyway). Suppose a and b have a common prime factor ℓ, then ℓ also has to divide $\mathcal{N}(\bar{\tau})^w = \mathcal{N}(\tau)^w = q^w$ (where $\mathcal{N}(\tau)$ denotes the norm of τ). Thus $\ell \mid q$. By using part (1) of this lemma we have $(\ell) = \mathfrak{p}\bar{\mathfrak{p}}$ as ideals over $\mathbb{Z}[\tau]$.

Let us assume for a moment that both $\mathfrak{p}$ and $\bar{\mathfrak{p}}$ divide (τ), then also both $\mathfrak{p}$ and $\bar{\mathfrak{p}}$ divide $(\bar{\tau})$, i.e., $\tau, \bar{\tau} \in \mathfrak{p}\bar{\mathfrak{p}} = (\ell)$. But this yields $\ell \mid \tau + \bar{\tau} = p$, a contradiction. Therefore let us assume now that $\mathfrak{p} \mid (\tau)$ and $\bar{\mathfrak{p}} \nmid (\tau)$. Since by assumption a and b have the common factor ℓ, they also have the common factor $\bar{\mathfrak{p}}$ from the ideal point of view, hence $\bar{\mathfrak{p}} \mid (\tau)^w$ if $a + b\tau = \tau^w$, a contradiction to the previous discussion. Similarly, we get the contradiction $\mathfrak{p} \mid (\bar{\tau})^w$ for the case $a + b\tau = \bar{\tau}^w$. $\square$

It is also important to know that no lattice points are on the "lower" edge of R_w. This result is also used to show the uniqueness of the digit set; see Proposition 12.2.

Lemma 4.3. *The following two statements hold.*

(1) *The only lattice point in $\Lambda_{\tau/2}$ lying on the line segment joining the points $v_0\tau^w$ and $v_1\tau^w$ is $\frac{1}{2}\tau^{w+1}$.*
(2) *The boundary of $\tau^w V$ has empty intersection with the lattice $\mathbb{Z}[\tau]$.*

Proof. We start by showing that there are no lattice points of $\Lambda_{1/2} = \langle \frac{1}{2}, \frac{\tau}{2} \rangle$ except $\frac{1}{2}\tau^w$ on the line going through $v_5\tau^w$ and $v_0\tau^w$. Every point on this line can be written as $\frac{1}{2}\tau^w + ti\tau^w$ with some real parameter t. Let us assume we have a point $\lambda = \frac{1}{2}\tau^w + a\frac{1}{2} + b\frac{\tau}{2}$ with a, $b \in \mathbb{Z}$ on this line. Furthermore, we may assume that $\gcd(a, b) = 1$ (as a minimality condition). We deduce $2ti\tau^w = a + b\tau$.

If $i \in \mathbb{Q}(\tau)$, then $4q - 1$ is a perfect square which is absurd since $4q - 1 \equiv -1 \mod 4$. Let us write $\tau^w = a_w + b_w\tau$, then Lemma 4.2 yields $\gcd(a_w, b_w) = 1$. Thus, and since $i \notin \mathbb{Q}(\tau)$, the only possible values for t are $\pm\frac{1}{2}$. Indeed $2ti\tau^w$ is an algebraic integer and therefore we have $2t \in \mathbb{Z}[\tau] \cap \mathbb{R} = \mathbb{Z}$ with $t \neq 0$ and $|t| < 1$. Now, we obtain a contradiction, since $\lambda = \frac{1}{2}(1 \pm i)\tau^w$ is not in $\Lambda_{1/2}$.

The results are now obtained as follows. Multiplying everything by τ yields the first statement of the lemma. Starting with $\frac{1}{2}\tau^{w+1} + ti\tau^{w+1}$ yields that there are no points from $\Lambda_{1/2}$ except $\frac{1}{2}\tau^{w+1}$ on the line joining $v_0\tau^w$ and $v_1\tau^w$. Note that this differs from the first statement of the lemma by the different lattice. The result for the third line (from $v_1\tau^w$ to $v_2\tau^w$) follows by taking conjugation and Lemma 4.2 with $\bar{\tau}^w = a_w + b_w\tau$. The remaining three sides of $\tau^w V$ follow by mirroring. ☐

We are also interested in the shortest vector in the lattice $\Lambda_{\tau/2}$.

Lemma 4.4. *The shortest non-zero vectors in the lattice $\Lambda_{\tau/2}$ are $\pm\tau/2$.*

Proof. First, let us note that $\left|\frac{\tau}{2}\right| = \frac{\sqrt{q}}{2}$. To find the shortest non-zero vector $a\frac{\tau}{2} + b\frac{\tau^2}{2} \in \Lambda_{\tau/2}$ we have to find all integers a and b such that $\left|a\frac{\tau}{2} + b\frac{\tau^2}{2}\right| \leq \frac{\sqrt{q}}{2}$. In particular we have to solve the inequality

$$|a + b\tau|^2 = \left(a + b\frac{p}{2}\right)^2 + b^2\left(q - \frac{1}{4}\right) \leq 1.$$

Obviously, if $b \neq 0$, then this inequality cannot be satisfied. Thus we may assume that $b = 0$. We obtain $a^2 \leq 1$, and the result follows. ☐

Part II. The Diophantine part

5. Overview

In this part of the article we show that the following proposition holds.

Proposition 5.1. *We use the set-up described in section 3 with the following restrictions. Suppose we are in one of the following cases:*

- *$q \in \{2, 4\}$ and $w \geq 7$,*
- *$q \in \{3\}$ and $w \geq 5$ or*
- *$5 \leq q \leq 500$ and $w \geq 4$.*

Then there exists a lattice point

$$a\tau + b(q - p\tau)$$

with $a \in \mathbb{Z}$, $q \nmid a$ and with $b \in \mathbb{Z}$ in the (open) rectangle R_w.

Since the proof of this proposition is long and technical, we start with an overview. In a nutshell, for fixed q, we can reduce the problem to checking only finitely many configurations w. Therefore the testing is possible algorithmically.

Let us look at an outline of the ideas used during the proof a bit. The resulting algorithm takes as input parameters q and p and returns a list of values for w which have to be investigated by other methods (i.e., yet no lattice point was found for these w). The details can be found in the last section of this part, section 10. In order to make this algorithm work, we have to check Proposition 5.1 for all but finitely many cases.

One major step to tackle this lemma is to reduce the lattice point problem into a Diophantine approximation problem. We show this in section 9. The existence result of the lattice points there is based on the theory of the geometry of numbers. More precisely, this gives us two lattice points (see Lemma 9.3) out of which we can construct a lattice point avoiding a smaller lattice (as it is required by Proposition 5.1); see Lemma 9.5. But, to make this work, we have to use linear independence of the two points, which is the challenging part during the proof.

We can reformulate this linear independence problem geometrically, which leaves us to show that there are no lattice points inside a certain smaller rectangle. To solve it, we bring this Diophantine approximation problem into a favorable form, which leaves us to show that the inequalities

$$\left|\log\left(\frac{a+b\tau}{|a+b\tau|}\right) - w\log\left(\frac{\tau}{|\tau|}\right) + k\frac{i\pi}{2}\right| < \chi q^{2-w/2} \tag{5.1a}$$

with $\chi = 9$ and

$$|a+b\tau| < \psi q^2 \tag{5.1b}$$

with $\psi = 4$ have no integer solutions.

With the previous inequality linear forms in logarithms come into play. The theory to solve this problem, unfortunately, provides only solutions for huge w (and fixed q); see section 7. The words "unfortunately" and "huge" here mean that it is not possible to test the remaining finitely many configurations in reasonable time. In order to reduce the bounds on possible integer solutions of (5.1) (and thus reducing the calculation time), we use a method due to Baker and Davenport [5] in section 8.

We are left with a bunch of small cases. Some remarks on how to check the lemma for these values directly can be found in section 6.

Note that the steps above were presented in reverse ordering (from the perspective, that we only use results, which were proven earlier in the article), since this is more the way one has to think when solving such a problem.

6. Testing directly

In this section, we collect some remarks on how to directly test whether Proposition 5.1 holds for fixed parameters. So let us fix q, p and w. We use the following criterion to find a lattice point $\lambda \in \Lambda_\tau \setminus \Lambda_{\tau^2}$ inside the rectangle R_w.

We first establish necessary and sufficient conditions for a complex number $z = x + iy$ to be contained in R_w.

Proposition 6.1. *Set*

$$B_1 = \frac{q^{w+1}}{4},$$
$$B_2 = \frac{q^{w+1}}{4} + \frac{q^{(w+1)/2}}{2}\sqrt{\frac{q-1/4}{q+2}},$$

and

$$B_3 = \frac{q^{w+1}}{2(4q-1)} - \frac{q^{w/2+1}}{\sqrt{4q-1}},$$

and let us write $\frac{1}{2}\tau^{w+1} = u_w + v_w\tau$. *Then* $\lambda = a\tau + b\tau^2 \in R_w$ *if and only if*

$$B_1 < a\left(\frac{u_w p}{2} + v_w q\right) + b\left(\frac{u_w}{2} - u_w q + \frac{v_w p q}{2}\right) < B_2 \tag{6.1a}$$

and

$$|a u_w + b(u_w p + v_w q)| < B_3. \tag{6.1b}$$

We can solve this system of inequalities and obtain finitely many pairs of integers (a, b). If we find a pair with $q \nmid a$, then Proposition 5.1 is true for this instance. Thus, Proposition 6.1 leads to a “searching algorithm” to solve Proposition 5.1 for a particular parameter set.

Proof. Let $\frac{1}{2}\tau^{w+1} = x_w + iy_w$. By elementary geometry we know that a point (x, y) that lies between the upper and lower length sides of R_w satisfies

$$\left(\frac{|\tau|^{w+1}}{2}\right)^2 < xx_w + yy_w < \frac{|\tau|^{w+1}}{2}\left(\frac{|\tau|^{w+1}}{2} + s\right).$$

Since

$$x_w + iy_w = u_w + \frac{v_w p}{2} + iv_w\sqrt{q - 1/4}$$

and since we want to have constraints for integers a and b in (x, y) with

$$x + iy = \lambda = a\tau + b\tau^2 = \frac{ap}{2} + \frac{b}{2} - bq + i(a + bp)\sqrt{q - 1/4},$$

we obtain

$$xx_w + yy_w = a\left(\frac{u_w p}{2} + \frac{v_w}{4}\right) + b\left(\frac{1}{2} - q\right)\left(u_w + \frac{v_w p}{2}\right) + (a + bp)v_w\left(q - \frac{1}{4}\right),$$

from which (6.1a) follows.

The inequality for (x, y) being in between the sides of R_w parallel to τ^w is given by

$$|xy_w - yx_w| < \frac{|\tau|^{w+1}}{2}\left(\frac{|v_0 - v_1|\,|\tau|^w}{2} - \sqrt{q}\right) = B_3\sqrt{q - 1/4}$$

and, likewise as above, we have

$$xy_w - yx_w = \left(a\frac{v_w p}{2} + bv_w\left(\frac{1}{2} - q\right) - (a + bp)\left(u_w + \frac{v_w p}{2}\right)\right)\sqrt{q - 1/4}.$$

The inequality (6.1b) follows.

Therefore the lattice point $\lambda \in \Lambda_\tau \cap R_w$ satisfies both inequalities stated in the lemma and, the other way around, all such points are inside R_w. $\square$

7. Huge bounds for w

This section deals with showing that the inequalities (5.1) have no integer solutions. We do this by providing a method to find for a fixed q all w such that (5.1) is satisfied. More precisely, we will give a (rather huge) bound on w such that solutions (if any) are only possible for smaller values.

However, for a single fixed q we still have (too) many possiblities to test all w; see Lemma 7.2 below and the text afterwards. Therefore we will reduce the upper bound of w by using a variant of the Baker–Davenport method [5] in section 8.

We can restrict ourselves to the following setting. We may assume $b > 0$ since otherwise $-a$, $-b$, and w would satisfy (5.1). Moreover, we may assume that $\gcd(a, b) = 1$. If a and b would have a common divisor d, then with a, b, and w also a/d, b/d, and w would satisfy (5.1).

As a first step we want to find a bound for w for a fixed integer $q \geq 2$.

Proposition 7.1. *For every q there exists an explicitly computable bound w_q such that the inequalities* (5.1) *do not have any integer solutions with $w \geq w_q$.*

The following lemma states the precise conditions when solutions of (5.1) are possible. Proposition 7.1 is a direct consequence of this result.

Lemma 7.2. *Solutions to* (5.1) *exist only if*

$$7.72 \cdot 10^{13} \log q \log(\sqrt{\psi} q) \log\left(4.87 w \frac{\max\{3\pi, 2\log q\}}{\log(\sqrt{\psi} q)}\right) > -\log\chi + \frac{w-4}{2}\log q \tag{7.1}$$

with $\chi = 9$ and $\psi = 4$ holds.

In particular, for

$$w \geq 8.68 \cdot 10^{15} \log q \log\log q \qquad \text{if } q \geq 13$$

and

$$w \geq 1.973 \cdot 10^{16} \qquad \text{if } q \in \{2, 3, \ldots, 12\}$$

the inequalities (5.1) *do not have any integer solutions.*

It is easy to see that for fixed q the inequality (7.1) cannot hold if w is large. For instance $q = 2$ yields $w \leq w_2 = 8.596 \cdot 10^{15}$ or for $q = 42$ we obtain $w \leq w_{42} = 2.747 \cdot 10^{16}$. This is one of the key results used in the proof of our main result, Theorem 2.3.

In the following, we denote by $h(\alpha)$ the *absolute logarithmic height*, which is defined as follows. Let α be an algebraic number of degree d and with minimal polynomial

$$a_0 \prod_{i=1}^{d} (X - \alpha_i),$$

then

$$h(\alpha) = \frac{1}{d}\Big(\log|a_0| + \sum_{i=1}^{d} \max\{0, \log|\alpha_i|\}\Big).$$

For the proof of Lemma 7.2 and in view of (5.1) we apply the following result due to Matveev [16].

Theorem 7.3 (Theorem 2.2 with $r = 1$ in [16])**.** *Denote by $\alpha_1, \ldots, \alpha_n$ algebraic numbers, not 0 or 1, by $\log\alpha_1, \ldots, \log\alpha_n$ determinations of their logarithms, by D the degree over $\mathbb{Q}$ of the number field $K = \mathbb{Q}(\alpha_1, \ldots, \alpha_n)$, and by $b_1, \ldots, b_n$ rational integers. Furthermore let $\kappa = 1$ if K is real and $\kappa = 2$ otherwise. For all integers j with $1 \le j \le n$ choose*

$$A_j \ge \max\{D\,h(\alpha_j), |\log\alpha_j|, 0.16\},$$

and set

$$B = \max\{1\} \cup \{|b_j|\,A_j/A_n : 1 \le j \le n\}.$$

Assume that

$$b_1\log\alpha_1 + \cdots + b_n\log\alpha_n \ne 0.$$

Then

$$\log|b_1\log\alpha_1 + \cdots + b_n\log\alpha_n| \ge -C(n,\kappa)\max\{1, n/6\}\,C_0W_0D^2\Omega$$

with

$$\Omega = A_1\cdots A_n,$$
$$C(n,\kappa) = \frac{16}{n!\,\kappa}e^n(2n+1+2\kappa)(n+2)(4(n+1))^{n+1}\left(\frac{1}{2}en\right)^{\kappa},$$
$$C_0 = \log\left(e^{4.4n+7}n^{5.5}D^2\log(eD)\right), \quad W_0 = \log(1.5eBD\log(eD)).$$

Proof of Lemma 7.2. We observe that in Matveev's theorem we have $n = 3$, $D = 4$, and $\kappa = 2$ since we use

$$\alpha_3 = \frac{a+b\tau}{|a+b\tau|}, \qquad \alpha_2 = \frac{\tau}{|\tau|}, \qquad \text{and} \qquad \alpha_3 = i.$$

Moreover, we set $b_3 = 1$, $b_2 = -w$ and $b_1 = k$.

Next, let us compute the heights of $\frac{a+b\tau}{|a+b\tau|}$ and $\frac{\tau}{|\tau|}$. Let us note that for an imaginary quadratic integer α the algebraic number $\alpha/|\alpha|$ is a zero of

$$|\alpha|^4(X - \alpha/|\alpha|)(X - \bar\alpha/|\alpha|)(X + \alpha/|\alpha|)(X + \bar\alpha/|\alpha|)$$

and therefore

$$h\left(\frac{\alpha}{|\alpha|}\right) = \frac{1}{4}(\log|a_0| + 4\log|\alpha/|\alpha||) \le \log|\alpha| \tag{7.2}$$

since $|a_0| \le |\alpha|^4$. We choose

$$\begin{aligned} A_3 &= 8\log(\sqrt{\psi}q) = 4\log(\psi q^2) \ge 4\log|a+b\tau|, \\ A_2 &= 2\log q = 4\log|\tau|, \\ A_1 &= 2\pi = 4|\log i|. \end{aligned}$$

Next we find an upper bound for k. Let us note that $\left|\log\frac{a+b\tau}{|a+b\tau|}\right| < 2\pi$ and from the consideration above we have

$$\left|\log\frac{\tau}{|\tau|}\right| \le \pi - \arctan(\sqrt{4q-1}) \le \frac{2\pi}{3}.$$

Therefore, a very crude estimate of inequality (5.1a) yields

$$\left|2\pi + w\frac{2\pi}{3} - k\frac{\pi}{2}\right| < \frac{\pi}{2},$$

and, thus, $k \le \frac{3w}{2}$. We choose

$$B = w\frac{\max\{3\pi, 2\log q\}}{A_3}.$$

Before we may apply Theorem 7.3 we have to check that our linear form in logarithms (i.e., the left hand side of (5.1a)) is non-zero. Let us assume for the moment the contrary. But assuming that the linear form in logarithms is zero expressed in geometric terms is that $\frac{1}{2}\tau^{w+1} + \frac{1}{2}(a\tau + b\tau^2)$ lies on the line segment joining the points $v_0\tau^w$ and $v_1\tau^w$, thus equals $\frac{1}{2}\tau^{w+1}$, which contradicts Lemma 4.3. In view of inequality (5.1a) and Theorem 7.3 we obtain (7.1).

We are left to compute the explicit bounds for w. Let us assume for the moment that $\max\{3\pi, 2\log q\} = 2\log q$, i.e., that $q \ge 112 > e^{3\pi/2}$. Let us note that under this assumption we have $\log(\sqrt{\psi}q) < 1.15\log q$. By a crude estimate we deduce that inequality (7.1) is not satisfied if

$$1.954 \cdot 10^{14} \log q \log w < w \tag{7.3}$$

holds, unless $w < 10^{10}$. Due to a result of Pethő and de Weger [20], namely their Lemma 2.2, an inequality of the form $A\log x \ge x$ with $A > e^2$ implies the inequality $x < 2A\log A$. Therefore we find an explicit bound for w, namely

$$w \ge 8.68 \cdot 10^{15} \log q \log\log q > 3.908 \cdot 10^{14} \log q \log\bigl(1.954 \cdot 10^{14} \log q\bigr),$$

which implies (7.3) and consequently the non-existence of solutions.

By solving inequality (7.1) for each integer $2 \le q < 112$ one can easily show that the explicit bounds stated in the lemma also hold for $q < 112$. □

Proof of Proposition 7.1. The result follows out of Lemma 7.2 since for fixed q Inequality (7.1) does not hold if w is sufficiently large. □

8. Reducing the bounds for w

The bounds from the previous section (Proposition 7.1) are too huge in order to test all remaining configurations in reasonable time. Now our aim is to reduce these bounds which is done below and works very well: For instance, the bound $w_2 = 8.596 \cdot 10^{15}$ is reduced to $\widetilde{w}_2 = 140$. We modify a method due to Baker and Davenport [5] to succeed; see Lemma 8.1 for details.

The remaining section deals with special cases and the occurrence of some linear dependence in the linear form in logarithms. Lemma 8.2 allows us to test for this linear dependence and in Lemma 8.3 we describe how to deal with this situation. At the end, we deal with two special cases (Lemma 8.4).

Let us denote the distance to the nearest integer by $\|\cdot\|$.

Lemma 8.1. *Suppose we have a bound w_q for w (i.e., inequalities* (5.1) *do not have any integer solutions for $w \ge w_q$). Fix a and b. Let P/Q be a convergent to*

$$\epsilon = \frac{2}{i\pi}\log\frac{\tau}{|\tau|}$$

with the properties that $\|Q\epsilon\| = \kappa/w_q$ for some $\kappa < 1/4$, but $\|Q\delta\| > 2\kappa$, where

$$\delta = \frac{2}{i\pi}\log\frac{a+b\tau}{|a+b\tau|}.$$

Then the inequalities (5.1) *do not have any integer solutions with our fixed a and b, and with*

$$w \geq \widetilde{w}_{a,b} = \frac{2}{\log q} \log\left(\frac{2\chi Q}{\kappa\pi}\right) + 4,$$

where $\chi = 9$.

Note that $\kappa < 1/4$ is formally not needed as an assumption in the lemma, but $\|Q\delta\| > 2\kappa$ implies this condition.

Proof of Lemma 8.1. Assume that $w_q > w \geq \widetilde{w}_{a,b}$ is satisfied and that we have a solution of (5.1). We multiply inequality (5.1a) by $2Q/\pi$ and use the notations of the lemma to obtain

$$|Q\delta - wQ\epsilon + k| < \frac{2\chi Q}{\pi} q^{2-w/2} \leq \kappa.$$

But on the other hand, we have

$$|Q\delta - wQ\epsilon + k| \geq \|Q\delta\| - w\,\|Q\epsilon\| > 2\kappa - w\frac{\kappa}{w_q} > \kappa.$$

Combining these two inequalities yields a contradiction. $\square$

We want to emphasize that Lemma 8.1 yields bounds only in the case when ϵ and δ are linearly independent over $\mathbb{Q}$. If this is not the case, then the considerations in the remaining section can be used. In particular, this is the case if $b = 0$ or $2a + bp = 0$ holds.

The following lemma allows us to test the linear dependence of $\log\bigl(\frac{\tau}{|\tau|}\bigr)$ and $\log\bigl(\frac{a+b\tau}{|a+b\tau|}\bigr)$ over $\mathbb{Q}$.

Lemma 8.2. *Suppose we have integers a and b such that $m = |a+b\tau|^2 < \psi^2 q^4$ with $\psi = 4$. Let us write $q = d_1 d_2^2$ and $m = m_1 m_2^2$, such that d_2 and m_2 are maximal with respect to* $\gcd(d_1, d_2) = \gcd(m_1, m_2) = 1$. *Set $m' = m_1/\gcd(m_1, d_1)$. With ν_ℓ being the ℓ-adic valuation, set $\alpha_\ell = \nu_\ell(q)$ for all primes $\ell \mid q$. For odd primes $\ell \mid q$ let $\alpha'_\ell = \alpha_\ell$ if $\ell \mid d_1 m'$ and put $\alpha'_\ell = \alpha_\ell/2$ otherwise. If $2 \mid q$, we put*

$$\alpha'_2 = \begin{cases} \alpha_2/2 & \text{if } d_1 \equiv m' \equiv 1 \mod 4,\\ \alpha_2 & \text{if } 2 \mid d_1 m' \text{ and if } d_1 \equiv 1 \mod 4 \text{ or } m' \equiv 1 \mod 4,\\ \alpha_2 & \text{if } 2 \nmid d_1 m' \text{ and if } d_1 \equiv 3 \mod 4 \text{ or } m' \equiv 3 \mod 4,\\ 2\alpha_2 & \text{if } 2 \mid d_1 m' \text{ and if } d_1 \equiv 3 \mod 4 \text{ or } m' \equiv 3 \mod 4. \end{cases}$$

Let N be the greatest common divisor of all α'_ℓ with primes $\ell \mid q$. Then $\log\bigl(\frac{\tau}{|\tau|}\bigr)$ and $\log\bigl(\frac{a+b\tau}{|a+b\tau|}\bigr)$ are linearly dependent over $\mathbb{Q}$ if and only if

$$\left(\frac{\tau}{|\tau|}\right)^{\eta} = \left(\frac{a+b\tau}{|a+b\tau|}\right)^{\vartheta} \tag{8.1}$$

for some positive integer $\vartheta \mid 24N$ and some integer η with $|\eta| < \vartheta\bigl(4 + \frac{2\log\psi}{\log q}\bigr)$.

Proof. It is immediate that $\log\bigl(\frac{\tau}{|\tau|}\bigr)$ and $\log\bigl(\frac{a+b\tau}{|a+b\tau|}\bigr)$ are linearly dependent over $\mathbb{Q}$ if and only if (8.1) holds. First, let us consider (8.1) as an equation in the ideal group of the field $K = \mathbb{Q}(\tau, \sqrt{d_1}, \sqrt{m'})$.

We aim to compute the prime ideal factorization of (ℓ) for every prime $\ell \mid q$. We already know by Lemma 4.2 that every such ideal (ℓ) splits in $\mathbb{Q}(\tau)$ and is therefore unramified. Furthermore, by definition d_1 and m' are coprime; therefore, (ℓ) with an odd prime $\ell \mid q$ is at most in one of the fields $\mathbb{Q}(\sqrt{d_1})$ and $\mathbb{Q}(\sqrt{m'})$ ramified. Hence, (ℓ) is ramified in K if and only if $\ell \mid d_1 m'$. Moreover, if the ideal (ℓ) is ramified, then the ramification index is exactly two. Altogether we get the following by using the fact that K is an Abelian extension of $\mathbb{Q}$.

- If $\ell \nmid d_1 m'$, then $(\ell) = \mathcal{J}_\ell \bar{\mathcal{J}}_\ell$, where $\mathcal{J}_\ell \mid (\tau)$ is the product of distinct prime ideals.
- If $\ell \mid d_1 m'$, then $(\ell) = \mathcal{J}_\ell^2 \bar{\mathcal{J}}_\ell^2$, where $\mathcal{J}_\ell \mid (\tau)$ is the product of distinct prime ideals.

Let us turn to the case $\ell = 2$. Recall that the ideal (2) ramifies in the quadratic field $\mathbb{Q}(\sqrt{d})$ if and only if $d \equiv 2 \mod 4$ or $d \equiv 3 \mod 4$. We have to distinguish between several cases.

- First, let us assume that the ideal (2) neither ramifies in $\mathbb{Q}(\sqrt{d_1})$ nor in $\mathbb{Q}(\sqrt{m'})$ (and consequently not in $\mathbb{Q}(\sqrt{d_1 m'})$ either), i.e., $d_1 \equiv m' \equiv 1 \mod 4$. Then (2) is also unramified in K and we have $(2) = \mathcal{J}_2 \bar{\mathcal{J}}_2$, where $\mathcal{J}_2$ is the product of distinct prime ideals.
- There is no situation, when the ideal (2) ramifies in exactly one of the fields $\mathbb{Q}(\sqrt{d_1})$, $\mathbb{Q}(\sqrt{m'})$ and $\mathbb{Q}(\sqrt{d_1 m'})$.
- Next, suppose the ideal (2) is ramified in exactly two of the fields $\mathbb{Q}(\sqrt{d_1})$, $\mathbb{Q}(\sqrt{m'})$ and $\mathbb{Q}(\sqrt{d_1 m'})$, i.e., one of the two cases
  - $2 \mid d_1 m'$ together with $d_1 \equiv 1 \mod 4$ or $m' \equiv 1 \mod 4$, or
  - $2 \nmid d_1 m'$ together with $d_1 \equiv 3 \mod 4$ or $m' \equiv 3 \mod 4$

  occurs. Then $(2) = \mathcal{J}_2^2 \bar{\mathcal{J}}_2^2$, where $\mathcal{J}_2$ is the product of distinct prime ideals.
- If the ideal (2) is ramified in all of the fields $\mathbb{Q}(\sqrt{d_1})$, $\mathbb{Q}(\sqrt{m'})$, and $\mathbb{Q}(\sqrt{d_1 m'})$, i.e., $2 \mid d_1 m'$ together with $d_1 \equiv 3 \mod 4$ or $m' \equiv 3 \mod 4$, then $(2) = \mathcal{J}_2^4 \bar{\mathcal{J}}_2^4$, where $\mathcal{J}_2$ is some prime ideal.

Therefore, by considering the definition of α'_ℓ, we obtain

$$(\tau) = \prod_{\substack{\ell \mid q \\ \ell \text{ prime}}} \mathcal{J}_\ell^{2\alpha'_\ell}.$$

Note that $\alpha'_\ell = \alpha_\ell/2$ only happens if α_ℓ is even. Furthermore we obtain

$$\left(\frac{\tau}{|\tau|}\right) = \prod_{\substack{\ell \mid q \\ \ell \text{ prime}}} \left(\frac{\mathcal{J}_\ell}{\bar{\mathcal{J}}_\ell}\right)^{\alpha'_\ell}.$$

Therefore $\left(\frac{\tau}{|\tau|}\right)$ is an Nth power if and only if N divides the greatest common divisor of all α'_ℓ with primes $\ell \mid q$.

Let us turn now from the ideal group point of view to the element point of view of equation (8.1). So far we have proved that $\log\left(\frac{\tau}{|\tau|}\right)$ and $\log\left(\frac{a+b\tau}{|a+b\tau|}\right)$ are linearly independent over $\mathbb{Q}$ if and only if there exist integers ϑ and η such that

$$\left(\frac{\tau}{|\tau|}\right)^{\eta} \left(\frac{a+b\tau}{|a+b\tau|}\right)^{-\vartheta} = 1$$

and $\vartheta \mid N$. Set $n = \gcd(\eta, \vartheta)$. By taking nth roots we obtain

$$\left(\frac{\tau}{|\tau|}\right)^{\eta/n} \left(\frac{a+b\tau}{|a+b\tau|}\right)^{-\vartheta/n} = \zeta_n, \tag{8.2}$$

where ζ_n is an nth root of unity. The group $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})$ is a subgroup of $\mathrm{Gal}(K/\mathbb{Q})$, and $\mathrm{Gal}(K/\mathbb{Q})$ is isomorphic to a subgroup of $(\mathbb{Z}/2\mathbb{Z})^3$. Since 24 is maximal with $\varphi(24) = 8$ (where φ is Euler's phi function), we deduce that $n \mid 24$. If we take equation (8.2) to the 24th power, we obtain the first statement of the lemma.

Because of (7.2) we have heights

$$h\left(\frac{\tau}{|\tau|}\right) = \frac{1}{2}\log q$$

and

$$h\left(\frac{a+b\tau}{|a+b\tau|}\right) \le \log|a+b\tau| = \log\sqrt{m} < \log\psi + 2\log q.$$

Comparing these heights on the left and right side of (8.1) we obtain

$$\frac{\eta}{2}\log q = \eta\, h\left(\frac{\tau}{|\tau|}\right) = \vartheta\, h\left(\frac{a+b\tau}{|a+b\tau|}\right) < \vartheta(\log\psi + 2\log q). \qquad \square$$

Lemma 8.3. *Suppose we have a bound w_q for w (i.e., inequalities (5.1) do not have any integer solutions for $w \ge w_q$), and suppose that we have for fixed a and b a linear dependence of the form*

$$\eta\log\left(\frac{\tau}{|\tau|}\right) = \vartheta\log\left(\frac{a+b\tau}{|a+b\tau|}\right),$$

such that $\vartheta > 0$.

(1) *Let P/Q be an expanded fraction of a convergent (i.e., $P/Q = P'/Q'$ for a convergent P'/Q') to $\epsilon = \frac{2}{\pi}\log\left(\frac{\tau}{|\tau|}\right)$ with the following properties. Suppose $W_Q = (Q+\eta)/\vartheta$ is largest possible with $W_Q < w_q$ (i.e., $Q < w_q\vartheta - \eta$) and $W_Q \in \mathbb{Z}$ such that*

$$Q\,|\epsilon - P/Q| < \frac{2\chi\vartheta}{\pi} q^{2-W_Q/2} \tag{8.3}$$

with $\chi = 9$ holds. If no such fraction P/Q exists, then set $W_Q = -\infty$.

(2) *Let W be the smallest positive integer such that the inequality*

$$\frac{2\chi\vartheta}{\pi} q^{2-W/2} \le \frac{1}{2(W\vartheta - \eta)} \tag{8.4}$$

with $\chi = 9$ holds.

Then the inequalities (5.1) do not have any integer solutions with our fixed a and b, and with

$$w \ge \widetilde{w}_{a,b} = \max\{W_Q + 1, W\}.$$

Proof. The assumption on the linear dependence yields an inequality of the form

$$\left|(w\vartheta - \eta)\log\left(\frac{\tau}{|\tau|}\right) - \vartheta k\frac{i\pi}{2}\right| < \chi\vartheta q^{2-w/2} \tag{8.5}$$

or with the notation of Lemma 8.1,

$$\left|\epsilon - \frac{\vartheta k}{w\vartheta - \eta}\right| < \frac{2\chi\vartheta}{(w\vartheta - \eta)\pi} q^{2-w/2}. \tag{8.6}$$

Note that due to a well-known theorem of Legendre we have the following: If

$$\frac{2\chi\vartheta}{(w\vartheta-\eta)\pi}q^{2-w/2}\leq\frac{1}{2(w\vartheta-\eta)^2},$$

which is true for large enough w, then $(\vartheta k)/(w\vartheta-\eta)=P'/Q'$, where P'/Q' is a convergent to ϵ. Since P' and Q' are coprime, we have $Q'\mid w\vartheta-\eta$, so $w\vartheta-\eta=Q$ for some multiple Q of Q'. $\square$

Lemma 8.4. *Suppose we have a bound w_q for w (i.e., inequalities* (5.1) *do not have any integer solutions for $w\geq w_q$) and suppose that $a=1$ and either $b=0$ or $b=-2p$. Let P/Q be an expanded fraction of a convergent (i.e., $P/Q=P'/Q'$ for a convergent P'/Q') to $\epsilon=\frac{2}{\pi}\log\left(\frac{\tau}{|\tau|}\right)$ with the following properties. Suppose $W_Q=Q$ is largest possible with $W_Q<w_q$ (i.e., $Q<w_q$) such that*

$$Q\left|\epsilon-P/Q\right|<\frac{2\chi}{\pi}q^{2-W_Q/2}$$

with $\chi=9$ holds. If no such convergent exists, then set $W_Q=-\infty$.

Then the inequalities (5.1) *do not have any integer solutions with our fixed a and b as above, and with*

$$w\geq\widetilde{w}_{a,b}=\max\{W_Q+1,25\}.$$

Note that the bound 25 is sharp for $q=2$, but for $q>2$ a better bound could be chosen.

Proof of Lemma 8.4. In both cases $\log\left(\frac{a+b\tau}{|a+b\tau|}\right)$ is an integral multiple of $\frac{i\pi}{2}$. Therefore, we consider the inequality

$$\left|w\log\left(\frac{\tau}{|\tau|}\right)-k\frac{i\pi}{2}\right|<\chi q^{2-w/2}\tag{8.7}$$

which is similar to (8.5). In the same way as before (proof of Lemma 8.3) and with the notation of Lemma 8.1 we obtain

$$\left|\epsilon-\frac{k}{w}\right|<\frac{2\chi}{w\pi}q^{2-w/2}$$

and that $\frac{k}{w}$ equals a convergent P'/Q' to ϵ if

$$\frac{2\chi}{w\pi}q^{(-w+4)/2}\leq\frac{1}{2w^2}\tag{8.8}$$

which is true for large w, in particular for all $w\geq 25$. Therefore a solution $w\geq 25$ to inequalities (5.1) corresponds to a fraction $P/Q=P'/Q'$ such that $Q=w$. $\square$

In order to get a reduced bound $\widetilde{w}_q$, we look at all possible combinations of a and b and calculate a bound $\widetilde{w}_{a,b}$ by the lemmata and considerations above. The bound $\widetilde{w}_q$ is the maximum of all these bounds.

9. Geometry of numbers

The theory of the geometry of numbers is used to show the existence of a lattice point in the rectangle R_w with the desired properties (i.e., out of the lattice Λ_τ but not in Λ_{τ^2}). We use two other rectangles inside R_w, one which is wide but low (called $\widetilde{R}_{\mathrm{WL}}$) and one which is narrow but high (called $\widetilde{R}_{\mathrm{NH}}$). Minkowski's lattice point theorem (Theorem 9.2) gives us the existence of a lattice point inside each of these two rectangles (see Lemma 9.3), and we are able to construct our

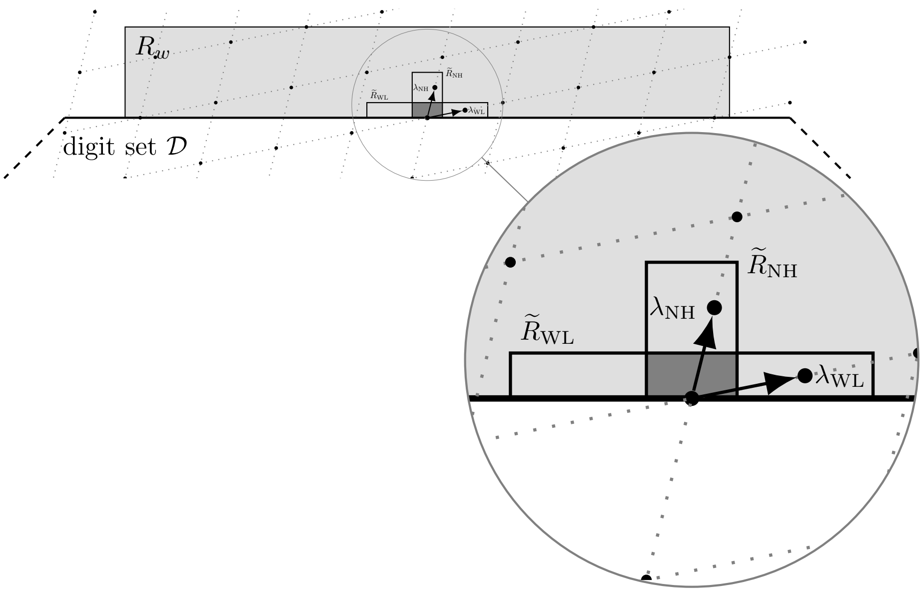

FIGURE 9.1. The rectangles $\widetilde{R}_{\mathrm{NH}}$ and $\widetilde{R}_{\mathrm{WL}}$, and points λ_{NH} and λ_{WL}.

desired lattice point out of it (Lemma 9.5), provided that the two found points are linearly independent. This is guaranteed if the intersection of the two mentioned rectangles with $\Lambda_{\tau/2}$ only contains $\tau^{w+1}/2$, which follows from the inequalities (5.1) by Lemma 9.1. So much for a short overview on this section; let us begin.

We use the lattices Λ_τ and $\Lambda_{\tau/2}$, which were defined in section 4. Throughout this section, we further use the rectangle $\widetilde{R}_{\mathrm{NH}}$ with vertices

$$\frac{\tau^{w+1}}{2} \pm i\frac{\tau^{w+1}}{|\tau|^{w+1}}q^2\sqrt{q+2}$$

and

$$\frac{\tau^{w+1}}{2} + \frac{\tau^{w+1}}{|\tau|^{w+1}}\frac{s}{4q} \pm i\frac{\tau^{w+1}}{|\tau|^{w+1}}q^2\sqrt{q+2}$$

with $s = \sqrt{\frac{q-1/4}{q+2}}$, and the rectangle $\widetilde{R}_{\mathrm{WL}}$ with vertices

$$\frac{\tau^{w+1}}{2} \pm i\frac{\tau^{w+1}}{|\tau|^{w+1}}\frac{1}{16q}\frac{q^{(w+1)/2}}{\sqrt{q-\frac{1}{4}}}$$

and

$$\frac{\tau^{w+1}}{2} + \frac{\tau^{w+1}}{|\tau|^{w+1}} \cdot 4q^{(3-w)/2}\left(q-\tfrac{1}{4}\right) \pm i\frac{\tau^{w+1}}{|\tau|^{w+1}}\frac{1}{16q}\frac{q^{(w+1)/2}}{\sqrt{q-\frac{1}{4}}}.$$

Note that these two rectangles are both contained in (the closure) of R_w. See also Figure 9.1.

In Lemmas 9.1, 9.4, and 9.5 we need that (at least) one of the conditions

(9.1)
- $w \geq 8$ and $q \geq 258$,
- $w \geq 9$ and $q \geq 17$,
- $w \geq 10$ and $q \geq 7$,
- $w \geq 11$ and $q \geq 5$,
- $w \geq 12$ and $q \geq 4$,
- $w \geq 13$ and $q \geq 3$, or
- $w \geq 16$ and $q \geq 2$

on q and w holds. These bounds are sharp in Lemma 9.4.

Lemma 9.1. *Suppose q and w satisfy conditions* (9.1)*. If inequalities* (5.1) *do not have any integer solutions (for given q and w), then the only lattice point of $\Lambda_{\tau/2}$ in $\widetilde{R}_{\mathrm{NH}} \cap \widetilde{R}_{\mathrm{WL}}$ is $\tau^{w+1}/2$.*

Since by construction $\widetilde{R}_{\mathrm{NH}} \cap \widetilde{R}_{\mathrm{WL}}$ is a rectangle with side lengths

$$2q^2\sqrt{q+2} \qquad \text{and} \qquad 4q^{(3-w)/2}\bigl(q-\tfrac{1}{4}\bigr), \tag{9.2}$$

therefore has an area which decreases with w, it seems very reasonable to assume that the only lattice point contained in $\widetilde{R}_{\mathrm{NH}} \cap \widetilde{R}_{\mathrm{WL}}$ is $\frac{1}{2}\tau^{w+1}$. In order to prove this result, we reformulate this geometric problem into a problem from Diophantine analysis (finding solutions for inequalities (5.1)).

Proof of Lemma 9.1. First let us note that the shortest vector of $\Lambda_{\tau/2}$ is $\frac{\tau}{2}$ which has length $\sqrt{q}/2$ (see Lemma 4.4). Therefore the angle between the lower long side of R_w and $\lambda \in \widetilde{R}_{\mathrm{NH}} \cap \widetilde{R}_{\mathrm{WL}}$ with $\lambda \neq \frac{\tau^{w+1}}{2}$ is at most

$$\arcsin \frac{4q^{(3-w)/2}\bigl(q-\frac{1}{4}\bigr)}{\frac{1}{2}\sqrt{q}} = \arcsin\bigl(8q^{2-w/2}\bigl(1-\tfrac{1}{4q}\bigr)\bigr) \tag{9.3}$$

in absolute values. Due to the conditions (9.1), the argument of the arcsine is less than 0.11 and we have $\arcsin(x) < 9x/8$; we obtain an upper bound for that angle. On the other hand, the angle between the vector τ^w and $a + b\tau$ is

$$\begin{aligned}\bigl|\arg a + b\tau - \arg \tau^w\bigr| &= \left|\log\left(\frac{a+b\tau}{|a+b\tau|}\right) - \log\left(\frac{\tau^w}{|\tau^w|}\right)\right| \\ &= \left|\log\left(\frac{a+b\tau}{|a+b\tau|}\right) - w\log\left(\frac{\tau}{|\tau|}\right) + k\frac{i\pi}{2}\right|\end{aligned}$$

for some integer k. This together with inequality (9.3) yields inequality (5.1a).

Now let us write $\lambda = \frac{1}{2}\tau^{w+1} + \frac{1}{2}(a\tau + b\tau^2)$. Further, by (9.2) we know that

$$\bigl|\tfrac{\tau}{2}(a+b\tau)\bigr|^2 \leq q^5\bigl(1+\tfrac{2}{q}\bigr) + 16q^{5-w}\bigl(1-\tfrac{1}{4q}\bigr)^2 < 4q^5$$

provided that $w \geq 4$. Thus

$$|a + b\tau| < \psi q^2$$

with $\psi = 4$ and a and b are bounded in terms of q. All together we obtain inequalities (5.1). ∎

In order to find at least one point inside each of the rectangles $\widetilde{R}_{\mathrm{NH}}$ and $\widetilde{R}_{\mathrm{WL}}$ we use Minkowski's lattice point theorem (for example, see Theorem II in Chapter III of Cassels [7]).

Theorem 9.2 (Minkowski's lattice point theorem)**.** *Let $S \subset \mathbb{R}^n$ be a compact point set of volume V which is symmetric about the origin and convex. Let Λ be any n-dimensional lattice with lattice constant $d(\Lambda)$. If $V \geq 2^n\, d(\Lambda)$, then there exists a pair of points $\pm\lambda \in \Lambda \cap S$, with $\lambda \neq 0$.*

Lemma 9.3. *For $j \in \{\mathrm{NH}, \mathrm{WL}\}$ there exists a lattice point $\lambda_j \in \Lambda_{\tau/2}$ in the rectangle $\widetilde{R}_j$.*

The situation of this lemma is shown in Figure 9.1.

Proof. First, note that the lattice $\Lambda_{\tau/2}$ has lattice constant

$$d\bigl(\Lambda_{\tau/2}\bigr) = \tfrac{1}{4}\, d(\Lambda_\tau) = \tfrac{1}{4} q\sqrt{q - \tfrac{1}{4}}.$$

Let us mirror the rectangle $\widetilde{R}_{\mathrm{NH}}$ on the line joining the points $v_0\tau^w$ and $v_1\tau^w$, and consider the rectangle $\widetilde{R}_{\mathrm{NH}}$ joint with the mirrored rectangle. We obtain a compact, symmetric around $\frac{1}{2}\tau^{w+1}$, convex set (a rectangle) of volume

$$2\frac{s}{4q} \cdot 2q^2\sqrt{q+2} = q\sqrt{q - \tfrac{1}{4}} = 4\, d\bigl(\Lambda_{\tau/2}\bigr)\,.$$

Now Minkowski's lattice point theorem yields a $\lambda_{\mathrm{NH}} \in \widetilde{R}_{\mathrm{NH}} \cap \Lambda_{\tau/2}$.

Let us construct λ_{WL} similarly: Again we mirror the rectangle $\widetilde{R}_{\mathrm{WL}}$ on the line joining the points $v_0\tau^w$ and $v_1\tau^w$ and consider the rectangle $\widetilde{R}_{\mathrm{WL}}$ joint with the mirrored rectangle. We obtain a compact, symmetric around $\frac{1}{2}\tau^{w+1}$, convex set again of volume

$$2 \cdot 4q^{(3-w)/2}\bigl(q - \tfrac{1}{4}\bigr) \cdot 2\frac{1}{8q}\frac{q^{(w+1)/2}}{\sqrt{q - \frac{1}{4}}} = q\sqrt{q - \tfrac{1}{4}} = 4\, d\bigl(\Lambda_{\tau/2}\bigr)\,.$$

Minkowski's lattice point theorem yields a $\lambda_{\mathrm{WL}} \in \widetilde{R}_{\mathrm{WL}} \cap \Lambda_{\tau/2}$. ☐

From now on we assume that we have λ_{NH} and λ_{WL} as in Lemma 9.3. The following result is needed in the proof of Lemma 9.5.

Lemma 9.4. *Suppose q and w satisfy conditions* (9.1)*. Then all lattice points of the form*

$$\frac{\tau^{w+1}}{2} + a\left(\lambda_{\mathrm{NH}} - \frac{\tau^{w+1}}{2}\right) + b\left(\lambda_{\mathrm{WL}} - \frac{\tau^{w+1}}{2}\right)$$

with non-negative integers a and b at most $2q$ are contained in the rectangle R_w.

Proof. All given points are contained in R_w if the two inequalities

$$2q \cdot q^2\sqrt{q+2} + 2q \cdot \frac{1}{16q}\frac{q^{(w+1)/2}}{\sqrt{q - \frac{1}{4}}} < \frac{1}{2}\sqrt{\frac{q^{w+1}}{4q-1}} - \sqrt{q}$$

and

$$2q \cdot \frac{s}{4q} + 2q \cdot 4q^{(3-w)/2}\bigl(q - \tfrac{1}{4}\bigr) < s$$

with $s = \sqrt{\frac{q-1/4}{q+2}}$ are satisfied. This is the case for the given conditions. ☐

With the construction above (and the assumptions of Lemma 9.4) we are in a position to prove the following lemma.

Lemma 9.5. *Suppose q and w satisfy conditions* (9.1). *If*

$$\widetilde{R}_{\mathrm{NH}} \cap \widetilde{R}_{\mathrm{WL}} \cap \Lambda_{\tau/2} = \left\{ \tfrac{1}{2}\tau^{w+1} \right\},$$

then there exists a $\lambda \in R_w$ with $\lambda \in \Lambda_\tau \setminus \Lambda_{\tau^2}$.

Proof. First we show that the lattice points $\mu_{\mathrm{NH}} = \lambda_{\mathrm{NH}} - \frac{1}{2}\tau^{w+1}$ and $\mu_{\mathrm{WL}} = \lambda_{\mathrm{WL}} - \frac{1}{2}\tau^{w+1}$ are linearly independent. Let us shift the origin to $\frac{1}{2}\tau^{w+1}$ and let us rotate the coordinate system such that the long "lower side" of the rectangle R_w, which contains the origin, is parallel to the real axis. In this new coordinate system, we write $\hat{\lambda}_{\mathrm{NH}}$ and $\hat{\lambda}_{\mathrm{WL}}$ for λ_{NH} and λ_{WL}, respectively. We have $\left|\mathrm{Re}\left(\hat{\lambda}_{\mathrm{NH}}\right)\right| < \left|\mathrm{Re}\left(\hat{\lambda}_{\mathrm{WL}}\right)\right|$ and $0 < \mathrm{Im}\left(\hat{\lambda}_{\mathrm{WL}}\right) < \mathrm{Im}\left(\hat{\lambda}_{\mathrm{NH}}\right)$, i.e., λ_{NH} and λ_{WL} are not colinear and therefore μ_{NH} and μ_{WL} are linearly independent.

Since we know now that μ_{NH} and μ_{WL} are linearly independent, there exists a basis ν_1, ν_2 of $\Lambda_{\tau/2}$ such that

$$\mu_{\mathrm{NH}} = \alpha_{11}\nu_1 \qquad \text{and} \qquad \mu_{\mathrm{WL}} = \alpha_{21}\nu_1 + \alpha_{22}\nu_2$$

with $0 < \alpha_{11}$ and $0 \leq \alpha_{21} < \alpha_{22}$. In our case this means there exists a basis ν_1, ν_2 for $\Lambda_{\tau/2}$ such that ν_1 and ν_2 are contained in the parallelogram with vertices 0, μ_{NH}, μ_{WL} and $\mu_{\mathrm{NH}} + \mu_{\mathrm{WL}}$. Moreover, by the assumptions of the lemma and Lemma 9.4 all lattice points of the form $\lambda = \frac{\tau^{w+1}}{2} + a\nu_1 + b\nu_2 \in \Lambda_{\tau/2}$ with $0 \leq a \leq 2q$ and $0 \leq b \leq 2q$ are contained in the rectangle R_w.

Now let us write

$$\nu_1 = \beta_{11}\tau + \beta_{12}\tau^2 \qquad \text{and} \qquad \nu_2 = \beta_{21}\tau + \beta_{22}\tau^2.$$

Since ν_1 and ν_2 as well as τ and τ^2 are bases to the same lattice Λ_τ we conclude that $\begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$. Our aim is to show that there exist non-negative integers a and b at most $2q$ such that $\lambda \in \Lambda_\tau$ but $\lambda \notin \Lambda_{\tau^2}$. Setting

$$\tfrac{1}{2}\tau^{w+1} = u_w \frac{\tau}{2} + v_w \frac{\tau^2}{2},$$

it suffices to prove that

$$\left(u_w + a\beta_{11} + b\beta_{12}\right)\frac{\tau}{2} + \left(v_w + a\beta_{21} + b\beta_{22}\right)\frac{\tau^2}{2} = \gamma_1\frac{\tau}{2} + \gamma_2\frac{\tau^2}{2}$$

for some $\gamma_1 \equiv 2 \mod 2q$ and $\gamma_2 \equiv 0 \mod 2q$ has a solution. But a solution can be found from a solution to the linear system

$$\begin{aligned} u_w + a\beta_{11} + b\beta_{12} &= 2, \\ v_w + a\beta_{21} + b\beta_{22} &= 0, \end{aligned}$$

modulo $2q$. Such a solution certainly exists since $\begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$. □

With the previous results, we are ready to prove the following proposition.

Proposition 9.6. *For every q there exists an explicitly computable bound w_q such that for each $w \geq w_q$ there exists a lattice point*

$$a\tau + b(q - p\tau)$$

with $a \in \mathbb{Z}$, $q \nmid a$ and with $b \in \mathbb{Z}$ in the rectangle R_w.

Proof. Proposition 7.1 states a similar result for the non-existence of solutions of inequalities (5.1). By Lemma 9.1 this translates to having the single intersection point $\frac{1}{2}\tau^{w+1}$ in the rectangles $\widetilde{R}_{\mathrm{NH}}$ and $\widetilde{R}_{\mathrm{WL}}$. This condition is then used in Lemma 9.5 to find a lattice point inside R_w as desired. □

10. An algorithm to test for fixed q

In order to prove Proposition 5.1 for given q and p, but all w—this means showing the existence of a lattice point in each rectangle R_w with the stated properties—we apply the following algorithm.[4]

Algorithm 10.1. *We fix q and fix a choice of $p \in \{-1, 1\}$ as input. This algorithm returns a list of values for w for which no lattice point in R_w exists. We proceed as follows.*

(1) *Compute an upper bound w_q for possible solutions (with $w < w_q$) using Matveev's Theorem* 7.3 *and in particular inequality* (7.1)*.*

(2) *Reduce the bound w_q by the Baker–Davenport method (section* 8*).*
   - (a) *Compute sufficiently many*[5] *(consecutive) convergents P/Q to $\epsilon = \frac{2}{i\pi}\log\frac{\tau}{|\tau|}$ and save them in a list $\mathcal{L}$. Precalculate and save κ with $\|Q\epsilon\| = \kappa/w_q$ as well.*
   - (b) *Use Lemma* 8.4 *to deal with the case $2a + bp = 0$ (i.e., $a = 1$, $b = -2p$) and compute the new bounds $\widetilde{w}_{a,b}$.*
   - (c) *For all integers a, b with $|a + b\tau| < \psi q^2$, $\psi = 4$ and with $b > 0$, coprime a, b and excluding the situations from step* (b) *do the following:*
     - (i) *Find η and ϑ such that*
$$\left(\frac{\tau}{|\tau|}\right)^{\eta} = \left(\frac{a + b\tau}{|a + b\tau|}\right)^{\vartheta}$$
using Lemma 8.2*.*
     - (ii) *If such η and ϑ do not exist in step* (i)*, find the convergents P/Q in $\mathcal{L}$ with smallest Q that satisfies $\kappa < \frac{1}{4}$ and $\|Q\delta\| > 2\kappa$ with $\delta = \frac{2}{i\pi}\log\frac{a+b\tau}{|a+b\tau|}$. Compute the new bound $\widetilde{w}_{a,b}$ due to Lemma* 8.1*.*
     - (iii) *If such η and ϑ in step* (i) *exist, find an expanded fraction P/Q of the convergents in $\mathcal{L}$ with largest Q that satisfies $W_Q = (Q+\eta)/\vartheta \leq w_q$ and* (8.4)*. Compute the new bound $\widetilde{w}_{a,b}$ due to Lemma* 8.3*.*
   - (d) *Calculate $\widetilde{w}_q$ as the maximum of all $\widetilde{w}_{a,b}$.*

(3) *For all $4 \leq w < \widetilde{w}_q$ (with exceptions obtained by taking into account the assumptions of Proposition* 5.1*), we verify Proposition* 5.1 *directly as described in section* 6*.*

Note that Algorithm 10.1 as it is written is not guaranteed to terminate.[6] The reason is that it might be impossible to find a convergent P/Q with the desired properties in step (ii). Stopping this search at some point and not using the reduced bound of step (2) will make the algorithm terminate for sure. However, a huge amount of w have to be checked in step (3) then.

---

[4]See `http://www.danielkrenn.at/koblitz2-non-optimal` for the code.

[5]"sufficiently many" means that in step (ii) of the algorithm a convergent can be found for all (non-dependent) situations a and b.

[6]One might call Algorithm 10.1 only a "procedure".

Proposition 10.2. *Let $q \geq 2$. If Algorithm* 10.1 *terminates, then it is correct, i.e., it returns a list of values for w for which no lattice point in R_w with the properties stated in Proposition* 5.1 *exists.*

Proof. Section 9 reduces the problem of finding lattice points in R_w to showing that inequalities (5.1) do not have any integer solutions. Step (1) provides a bound for w; it can be computed effectively according to Proposition 7.1. Step (2) reduces this bound. We get a bound for each possible combination of a and b (all the different cases are analyzed in section 8); correctly determining whether we have a linear dependence is done via Lemma 8.2.

Taking the maximum of all these bounds yields $\widetilde{w}_q$, a new bound. In step (3) all remaining w are checked by a direct search according to Proposition 6.1. Since this proposition finds a lattice point if and only if one exists, we are able to classify all the w and return a list of the exceptional values. ☐

Proof of Proposition 5.1. We apply Algorithm 10.1. ☐

Part III. The part with the digits

11. Overview

In this part of the article, we construct the actual counterexamples to the minimality of the width-w non-adjacent forms (see section 2 for the relevant definitions). This means we have to find an expansion of a lattice point with a lower number of non-zero digits than the width-w non-adjacent form (w-NAF) of this point.

We reuse the ideas of Heuberger and Krenn [12] for our construction. This work also tells us that (if it exists) a counterexample using a (multi-)expansion of weight two can be found. Therefore, we will try to find

$$A\tau^{w-1} + B = C\tau^{w} + D = E\tau^{2w} + F\tau^{w} + G, \tag{11.1}$$

where the most left and the most right parts of the equation are valid digit expansions, i.e., A, B, E, F and G are digits. Moreover, we assume that D is a digit (equal to G), but, in order to get a counterexample to minimality, C is not allowed to be a digit. However, the point C is important during the construction of this counterexample: we will have $C = E\tau^{w} + F$ and $D = G$, and, more important, there will be a change Δ with $\tau C = A + \Delta$ and $D = B - \Delta\tau^{w-1}$.

Some explicit constructions are given in section 13 and Proposition 17.1, but most of the time we will consider a more general situation. There, the existence of a construction as above relies on Proposition 5.1, which was proven in the previous part. This lemma gives us the point C. The change Δ is discussed in section 14, and section 15 deals with the digits B, D, and G. Everything is glued together in sections 16 and 17.

We begin with a section which deals with the digit set we use. Note that this digit set is strongly related to the Voronoi cell defined in section 3.

12. Digit sets

In this section, we make a formal definition of the used digit set. This is equivalent to the definition stated in section 2 but uses the Voronoi cell to model the minimal norm property. Afterwards we show that this choice of digits is unique.

Definition 12.1 (Minimal norm digit set)**.** Let w be an integer with $w \geq 2$ and $\mathcal{D} \subseteq \mathbb{Z}[\tau]$ consist of 0 and exactly one representative of each residue class of $\mathbb{Z}[\tau]$ modulo τ^w that is not divisible by τ. If all such representatives $\eta \in \mathcal{D}$ satisfy $\eta \in \tau^w V$, then $\mathcal{D}$ is called the *minimal norm digit set modulo* τ^w.

The minimal norm digit set above is uniquely determined; see below.

Proposition 12.2. *Let $\mathcal{D}$ be a minimal norm digit set modulo τ^w. Then $\mathcal{D}$ is uniquely determined. In particular, there exists a unique element of minimal norm in each residue class modulo τ^w which is not divisible by τ.*

This proposition was proved for $q = 2$ in Avanzi, Heuberger and Prodinger [4]. The proof there uses a result of Meier and Staffelbach [17], namely their Lemma 2. This lemma and the result for $q = 2$ can be generalized in a straightforward way for arbitrary primes q. We use a different method in this article, which gives us the result for arbitrary integers q here.

Proof of Proposition 12.2. Digits strictly inside the scaled Voronoi cell $\tau^w V$ are unique, since they are closer to 0 than to any other point of $\tau^w\mathbb{Z}[\tau]$ by the definition of the Voronoi cell. In Lemma 4.3 we have already shown that there are no lattice points on the boundary of the scaled Voronoi cell $\tau^w V$. Therefore no non-uniqueness can occur and thus the proposition is proved. $\square$

13. Non-optimality for some values of w

We start here with a first family of counterexamples to optimality. We show the existence of expansions as in (11.1). The following propositions are devoted to the case when $w = 2$, where we give an explicit construction. Afterwards, we consider the case $w = 3$.

Proposition 13.1. *Let $q \geq 3$ and $p = 1$, and set $a = \lceil q/2 \rceil$. Set*

$$\begin{aligned} & & E &= 1, \\ A &= (1-a)\tau + a - q, & F &= q - a, \\ B &= 1 - \tau, & G &= (a-1)\tau + 1. \end{aligned}$$

Then

$$A\tau + B = E\tau^4 + F\tau^2 + G \tag{13.1}$$

and both sides of the equation are valid digit expansions, i.e., the 2*-NAF is not a minimal digit expansion.*

Proposition 13.2. *Let $q \geq 3$ and $p = -1$, and set $a = \lceil q/2 \rceil$. Set*

$$\begin{aligned} & & E &= 1, \\ A &= (1-a)\tau - a + q, & F &= q - a, \\ B &= (a-1)\tau - 1, & G &= -\tau - 1. \end{aligned}$$

Then

$$A\tau + B = E\tau^4 + F\tau^2 + G \tag{13.2}$$

and both sides of the equation are valid digit expansions, i.e., the 2*-NAF is not a minimal digit expansion.*

Proof of Propositions 13.1 *and* 13.2. This proof is assisted[7] by SageMath [24]. Equality in (13.1) and (13.2) is easy to verify and can be done by a simple symbolic calculation over the ring $\mathbb{Z}[\tau]$.

We are left with checking that we have valid digit expansions, i.e., that all claimed digits are indeed digits. We do this by showing that these quantities are closer to 0 than to any neighboring lattice point of $\tau^2\mathbb{Z}[\tau]$ (see also the construction of the digit set via Voronoi cells, section 12). These neighboring lattice points are exactly the τ^2-multiples of the points 1, τ, $\tau - p$, -1, $-\tau$ and $-\tau + p$. This leads to six inequalities for each digit. Note that we have $|a + b\tau|^2 = a^2 + b^2 q + pab$. We also check that the point $C = \tau^2 + (q - a) = p\tau - a$ is not a digit for technical reasons, which leads to one additional inequality. Note that we have $\Delta = -pa$ (with the notation of section 11) here.

However, distinguishing between $p = 1$ and $p = -1$ and between $q = 2\widetilde{q}$ (even) and $q = 2\widetilde{q} - 1$ (odd), we get 25 polynomials (each as difference of the two sides of an inequality) out of $\mathbb{Z}[\widetilde{q}]$. All these polynomials have degree at most 3 and a positive leading coefficient, and we can show, by using interval-arithmetic, that all their roots are smaller than 2. This means all polynomials are positive for $\widetilde{q} \geq 2$ and therefore the inequalities are satisfied.

Since the constant terms of the claimed digits are not divisible by q, they themselves are not divisible by τ. Therefore, we get valid digit expansions, which finishes the proof. $\square$

Proposition 13.3. *Let $q \geq 3$ and $p = 1$, and set $a = \lceil q/2 \rceil$ and $b = \lceil q^2/(6q - 2) \rceil$. For odd q set*

$$\begin{aligned} & & E &= 1, \\ A &= (1 - q)\tau + (a - 1)q - 1, & F &= (q - a)\tau + q - a, \\ B &= (q - a)\tau + a - 1, & G &= -a\tau + a - 1, \end{aligned}$$

and for even q set

$$\begin{aligned} & & E &= 1, \\ A &= (-a - b)\tau + (a - 1)q + 1, & F &= (q - a - 1)\tau + q - b, \\ B &= a\tau - aq + 1, & G &= -(a - 1)\tau + (a - 1)q + 1. \end{aligned}$$

Then

$$A\tau^2 + B = E\tau^6 + F\tau^3 + G$$

and both sides of the equation are valid digit expansions, i.e., the 3-NAF is not a minimal digit expansion.

Proposition 13.4. *Let $q \geq 3$ and $p = -1$, and set $a = \lceil q/2 \rceil$ and $b = \lceil q^2/(6q - 2) \rceil$. For odd q set*

$$\begin{aligned} & & E &= 1, \\ A &= (q - 1)\tau + (a - 1)q - 1, & F &= (q - a)\tau + q - a, \\ B &= -a\tau + 1 - a, & G &= (q - a)\tau + 1 - a, \end{aligned}$$

[7]The worksheet can be found at `http://www.danielkrenn.at/koblitz2-non-optimal`.

and for even q set

$$\begin{aligned} & & E &= 1, \\ A &= (a+b)\tau + (a-1)q + 1, & F &= (q-a-1)\tau - q + b, \\ B &= -a\tau - aq + 1, & G &= (a-)\tau + (a-1)q + 1. \end{aligned}$$

Then

$$A\tau^2 + B = E\tau^6 + F\tau^3 + G$$

and both sides of the equation are valid digit expansions, i.e., the 3*-NAF is not a minimal digit expansion.*

Proof of Propositions 13.3 *and* 13.4*.* We use the same machinery as in the proof of Propositions 13.1 and 13.2.

If $p = 1$ and q is odd, then we take $C = (1-a)\tau - a$ for our technical point. We use $\Delta = 1 - \tau$. Note that $B = a - 1 + (q-a)\tau$ and $D = G = a - 1 - a\tau$. To verify that A, B, E, F and G are digits, we, again, calculate the distance to 0 and its neighbors in $\tau^3\mathbb{Z}[\tau]$. This results in 31 inequalities, which are of polynomial type. This leaves us to check if elements out of $\mathbb{Z}[\widetilde{q}]$ with degree at most 4 and $q = 2\widetilde{q} - 1$ are positive, see the proof above for details. We can affirm this (it was done algorithmically).

For $p = -1$ and odd q we use $C = (1-a)\tau + a$, $\Delta = -\tau - 1$, $B = 1 - a - a\tau$, $D = G = 1 - a + (q-a)\tau$ and proceed in the same manner.

In the case that $p = 1$ and q is even, we have to be more careful because of the definition of b. We start similarly and take $C = -a\tau - b$ for our technical point and use $\Delta = q - 1$. Moreover, we use $B = 1 + a\tau^2$ and $D = G = 1 - (a-1)\tau^2$. The verification of the digits is done as above, but we take b into account.

Since $b = \lceil q^2/(6q-2) \rceil$ is obviously not polynomial, we cannot expect that these distance inequalities are polynomials. To deal with the ceil-rounding, we use for the moment $b_1 = q^2/(6q-2) + 1$ and $C_1 = -a\tau - b_1$ (note, this is not a lattice point) instead of b and C in the resulting inequalities; we will correct this later. As this b_1 is rational, we multiply the inequalities first by $6q - 2$ and then check whether the resulting polynomials (difference of the two sides of the inequality) out of $\mathbb{Z}[\widetilde{q}]$ are positive for all $q = 2\widetilde{q}$ as in the proof above. This verification is successful. This particularly means, that the point $-a\tau - b_1$ is not in $\tau^3 V$ (the scaled Voronoi cell containing the digits).

Next, consider $C_0 = -a\tau - b_0$ with $b_0 = q^2/(6q-2)$. We have

$$C_0 = \frac{\tau^3}{2} + ci\tau^3$$

with $c = \operatorname{Im}(\tau)/(3q-1)$. This means that C_0 is on the boundary of $\tau^3 V$ (as $\tau^3/2$ is on the boundary). The point $C = -a\tau - b$ is located on the line from C_0 to C_1. Due to convexity of $\tau^3 V$, this lattice point C lies on the outside of $\tau^3 V$ and, thus, is not a digit.

If we have $p = -1$, still with an even q, the proof works similarly, but we use $C = -a\tau + b$. □

14. Existence of a small change

From now on, in contrast to the previous section, where we have given an explicit construction for a counterexample to minimality of the w-NAF, we start with a

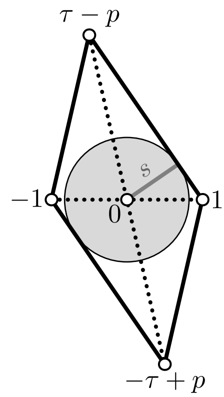

FIGURE 14.1. Parallelogram P of Lemma 14.1 for $q = 5$ and $p = 1$.

different approach. It still builds up on the ideas mentioned in the introduction of Part III and will be described fully in section 16. Before we are ready for this alternative construction, we need a couple of auxiliary results. In this section, we look at the change Δ a bit more closely (see Lemma 14.2), but let us start with the following lemma.

Lemma 14.1. *Let P be the parallelogram with vertices 1, $\tau - p$, -1, $-\tau + p$. Then a disc with center 0 and radius $s = \sqrt{\frac{q-1/4}{q+2}}$ fits exactly (i.e., the radius s is largest possible) in the parallelogram P.*

The situation is shown in Figure 14.1. Note that we have $s \geq \sqrt{7}/4$.

Proof. We can assume $p = 1$, since the other situation ($p = -1$) is just mirrored. First, we calculate the difference of the areas of the two triangles with vertices $\tau - 1$, $-\frac{1}{2}$, 1 and $\tau - 1$, $-\frac{1}{2}$, 0 and get

$$\tfrac{1}{2}\operatorname{Im}(\tau)\left(\tfrac{3}{2} - \tfrac{1}{2}\right) = \tfrac{1}{2}\sqrt{q - \tfrac{1}{4}}.$$

This area is equal to the area of the triangle with vertices 0, 1 and $\tau - 1$, which is

$$\tfrac{1}{2}s\sqrt{\operatorname{Im}(\tau)^2 + \tfrac{9}{4}} = \tfrac{1}{2}s\sqrt{q+2}.$$

Therefore the desired s follows. □

Lemma 14.2. *Let z be in the rectangle R_w. Then there exists a*

$$\Delta \in \{-1, 1, \tau - p, -\tau + p\},$$

such that $z - \Delta$ is in the interior of the scaled Voronoi cell $\tau^w V$.

Note that we will have $A = z - \Delta$ and $z = \tau C$ in the construction of our expansions. Moreover, the point z is the point inside the rectangle R_w whose existence was shown as the main result of Part II, Proposition 5.1.

Proof of Lemma 14.2. Let P_z be the parallelogram with vertices $z + 1$, $z + \tau - p$, $z - 1$, $z - \tau + p$, i.e., a shifted version of the parallelogram P of Lemma 14.1; see also Figure 14.1. Since z is in the (open) rectangle R_w, its distance to the line L from $\tau^w v_0$ to $\tau^w v_1$ is smaller than s (note that s is the height of the rectangle R_w, section 3). Further, since the rectangle R_w starts $\sqrt{q}$ away from the points $\tau^w v_0$ and $\tau^w v_1$, respectively, the distance from z to one of those two points is larger than $\sqrt{q}$. Therefore, the line L cuts the parallelogram P_z into two parts. The two cutting points are on different edges of P_z. This means, that there exists a vertex of the parallelogram P_z on that side of the line L, where there is no rectangle R_w. Such a point lies in the interior of the Voronoi cell $\tau^w V$, which can be seen using some properties of $\tau^w V$, including that two neighboring edges of $\tau^w V$ have an obtuse angle at their point of intersection and that a disc with center 0 and radius $\sqrt{q}$ is contained in $\tau^w V$. $\square$

15. Change in least significant digit

For our construction of the counterexample, we also have to deal with the change in the least significant digit, i.e., with the digits B, D and G in the expansions (11.1).

Lemma 15.1. *We get*

$$d_A = \frac{\sqrt{q}}{2} - \frac{1}{4\sqrt{q}} \quad \text{and} \quad d_B = \frac{1}{4\sqrt{q}}$$

in Figure 15.1.

Proof. The triangle $v_{-1/2}\alpha v_0$ is similar to the triangle $0 v_{1/2} v_1$. Therefore

$$d_A = \operatorname{Im}(v_0)\frac{\frac{1}{2}|\tau|}{\operatorname{Im}(v_1)} = \frac{q-\frac{1}{2}}{2q}\sqrt{q} = \frac{\sqrt{q}}{2} - \frac{1}{4\sqrt{q}}.$$

To calculate d_B we start as above. The triangle $v_{3/2}\beta v_1$ is similar to the triangle $-\overline{\tau}\delta 0$, so $d_B = d_D d_C / \left|-\overline{\tau}\right|$. We have

$$d_C = \frac{1}{2}\sqrt{(\operatorname{Im}(v_1) - \operatorname{Im}(v_2))^2 + \left(\frac{1}{2}\right)^2} = \frac{1}{2}\sqrt{\frac{1}{4} + \frac{1}{16(q-\frac{1}{4})}} = \frac{1}{4}\sqrt{1 + \frac{1}{4q-1}}$$

by the Pythagorean theorem. The distance d_D is the projection of $-\overline{\tau}$ on the normalized vector with direction $i\tau$. Therefore

$$d_D = \operatorname{Re}\left(\overline{-\overline{\tau}} i \frac{\tau}{|\tau|}\right) = -\frac{1}{|\tau|}\operatorname{Re}\bigl(i\tau^2\bigr) = \frac{1}{|\tau|}\operatorname{Im}\bigl(\tau^2\bigr) = \frac{1}{\sqrt{q}}\frac{1}{2}\sqrt{4q-1} = \sqrt{1 - \frac{1}{4q}}.$$

Now we can calculate d_B as

$$d_B = \frac{1}{\sqrt{q}}\sqrt{1 - \frac{1}{4q}}\frac{1}{4}\sqrt{1 + \frac{1}{4q-1}} = \frac{1}{4\sqrt{q}}. \qquad \square$$

Lemma 15.2. *For each point in $z \in \mathbb{C}$ there is a lattice point $u \in \mathbb{Z}[\tau]$ not divisible by τ with*

$$|z - u| < \tfrac{1}{2}|\tau| + 1 = \tfrac{1}{2}\sqrt{q} + 1$$

Proof. First note that if $\eta \in \mathbb{Z}[\tau]$ is divisible by τ, then $\eta + 1$ and $\eta - 1$ are not divisible by τ. Consider the lines $z(z + \tau/2)$ and $z(z - \tau/2)$. One of these lines cuts a horizontal line with lattice points $\eta_k = j\tau + k$, for some fixed $j \in \mathbb{Z}$ and all $k \in \mathbb{Z}$, on it. This means that a lattice point u can be found by first going from z at most a distance of $|\tau|/2$ and then at most 1 on the horizontal line. Strictly smaller holds since both directions are linearly independent. $\square$

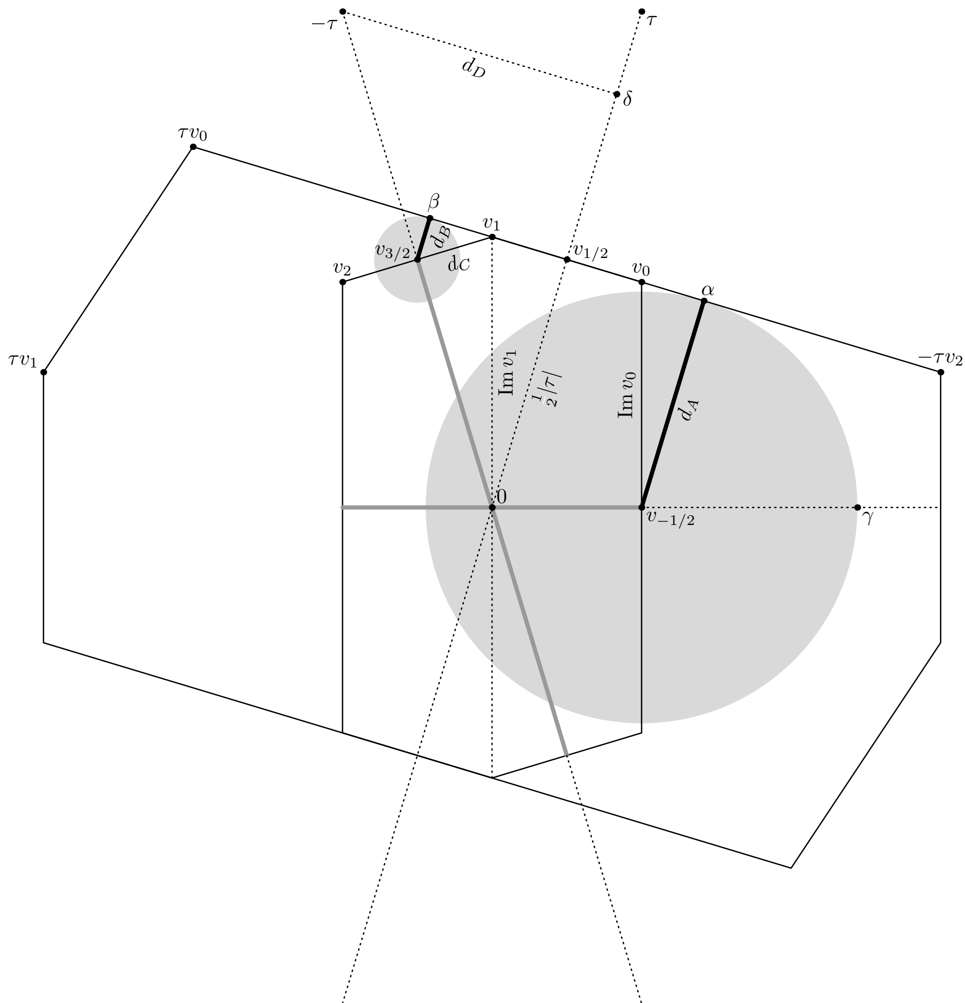

FIGURE 15.1. Distances and points used in Lemma 15.1 and Proposition 15.3.

Proposition 15.3. *If either $w \geq 4$ and $q \geq 11$ or $w \geq 8$, then there is a possible compensate change.*

More precisely, if $w \geq 5$ or $q \geq 5$, then there is a digit a such that $a + \tau^{w-1}$ is a digit as well. If either $w \geq 4$ and $q \geq 11$ or $w \geq 8$, then there is a digit b such that $b + \overline{\tau}\tau^{w-1}$ is a digit as well.

Proof. The digit set $\mathcal{D}$ is contained in $\tau^w V$. Consider the line form $\tau^{w-1} v_{-1/2} = \frac{1}{2}\tau^{w-1}$ to $-\tau^{w-1} v_{-1/2} = -\frac{1}{2}\tau^{w-1}$. From each end point of that line there is a lattice point not divisible by τ within a radius $\sqrt{q}/2 + 1$ by Lemma 15.2. With the quantities of Lemma 15.1 one can easily check that the inequality

$$\frac{1}{2}\sqrt{q} + 1 < |\tau|^w \frac{d_A}{|\tau|} = \sqrt{q}^{w-1}\left(\frac{\sqrt{q}}{2} - \frac{1}{4\sqrt{q}}\right)$$

holds for either $w \geq 5$ or $q \geq 5$ (fixing $w = 2$ makes it easy to check the inequality for $q \geq 5$, then use monotonicity in w; use the same argumentation starting with $q = 2$ and $w \geq 5$). Thus, the lattice points found above are in the interior of $\tau^w V$, and so are our desired digits a and $a + \tau^{w-1}$.

Now we do similarly to get digits b and $b + \overline{\tau}\tau^{w-1}$. We consider the line from $\tau^{w-1} v_{3/2} = -\frac{\overline{\tau}}{2}\tau^{w-1}$ to $-\tau^{w-1} v_{3/2} = -\frac{\overline{\tau}}{2}\tau^{w-1}$ and have to check the inequality

$$\frac{1}{2}\sqrt{q} + 1 < |\tau|^w \frac{d_B}{|\tau|} = \sqrt{q}^{w-1} \frac{1}{4\sqrt{q}}.$$

That inequality is satisfied either if $w \geq 4$ and $q \geq 11$ or if $w \geq 8$, which can again be checked easily by monotonicity arguments. For $w = 2$ or $w = 3$ the inequality is never satisfied. $\square$

16. Finding a non-optimal w-NAF

In contrast to section 13, where we have given an explicit construction for a counterexample to minimality of the w-NAF, we use a different approach here. It still builds up on the ideas mentioned in the introduction of Part III, i.e., for our construction we consider an element

$$z = A\tau^{w-1} + B \in \mathbb{Z}[\tau]$$

with non-zero digits $A \in \mathcal{D}$ and $B \in \mathcal{D}$ with the following additional properties. We want to find a change Δ with

- $\tau \mid (A + \Delta)$,
- $\tau^2 \nmid (A + \Delta)$,
- $\tau^{-1}(A + \Delta) \notin \mathcal{D}$, and
- $B - \Delta\tau^{w-1} \in \mathcal{D}$.

We will restrict ourselves here to

$$\Delta \in \{-1, 1, \tau - p, -\tau + p\},$$

which turns out to be a good choice. (Note that this restriction was already used in section 14; we only will relax it in Proposition 17.1)

Then the w-NAF-expansion of $C = \tau^{-1}(A + \Delta)$ has weight at least 2, because it (its value) is not a digit. We obtain

$$z = A\tau^{w-1} + B = C\tau^w + \left(B - \Delta\tau^{w-1}\right),$$

which shows that z has a (non-w-NAF) expansion with weight 2 and a w-NAF-expansion with weight at least 3. For all of our cases, the right hand side of the previous equation can be rewritten in an expansion

$$z = E\tau^{2w} + F\tau^w + G$$

with some digits E, F and $G = D = B - \Delta\tau^{w-1}$.

The finding of a point $A + \Delta$ is based on the main result of Part II, A and Δ are discussed in section 14, and section 15 is devoted to get digits B and $G = D$. We just have to glue all the results together, which is done in the proposition below and in the next section. Alternatively, a direct search can be used to find those lattice point configurations; we use this when Part II does not provide a result.

Proposition 16.1. *Suppose either $w \geq 4$ and $q \geq 11$ or $w \geq 8$ (as in Proposition 15.3). Moreover, set*

$$C = \tau^{-1}\left(a\tau + b(q - p\tau)\right)$$

and suppose we have

$$\tau C \in R_w$$

for some $a \in \mathbb{Z}$ with $q \nmid a$ and $b \in \mathbb{Z}$. Then there exist digits A, B and D such that

$$A\tau^{w-1} + B = C\tau^w + D.$$

Note that C cannot be a digit because of the following reasons. The point τC lies in the rectangle R_w, and thus C is outside of $\tau^w V$; see also Remark 3.1. Since C not divisible by τ (because $q \nmid \ell$) either, it has an expansion of weight at least 2. Using this expansion leads to our desired counterexample.

Proof. Consider the lattice point $\tau C \in R_w$. By Lemma 14.2 there exists a $\Delta \in \{-1, 1, \tau - p, -\tau + p\}$ such that

$$A = \tau C - \Delta$$

lies in the interior of the scaled Voronoi cell $\tau^w V$. Since C is a lattice point (and Δ is not divisible by τ), the lattice point A is not divisible by τ. Therefore A is a digit (see section 12).

Proposition 15.3 gives us the digits B and $D = B - \Delta\tau^{w-1}$. This completes the proof. □

17. Collecting all results

In this final section, we prove Theorems 2.3 and 2.4 and Proposition 2.5.

Proof of Theorem 2.3. If we have a lattice point $\tau C \in R_w$ not divisible by τ^2, then we are able to construct a counterexample by Proposition 16.1. Fortunately Proposition 9.6 provides the explicitly computable bound w_q and that for all $w \geq w_q$ a τC as above exists. □

Before proving Theorem 2.4 we need to consider one special case first.

Proposition 17.1. *Let $q = 4$ and $p \in \{-1, 1\}$ and $w = 6$. Set*

$$\begin{aligned} & & E &= 1, \\ A &= 7\tau - 66p, & F &= -16p\tau + 10, \\ B &= -65, & G &= 10p\tau - 9. \end{aligned}$$

Then

$$A\tau^5 + B = E\tau^{12} + F\tau^6 + G \tag{17.1}$$

and both sides of the equation are valid digit expansions, i.e., the 6-NAF is not a minimal digit expansion.

Proof. We use a direct search following the same ideas as presented above (especially in section 16). However, we have to relax our conditions on Δ, in particular, we use $\Delta = -2p$. Moreover, we have $C = 17p\tau - 10$ as intermediate result in the construction. □

Proof of Proposition 2.5 *and Theorem* 2.4. We start as in the proof of Theorem 2.3, i.e., we need a lattice point $\tau C \in R_w$ not divisible by τ^2. The main result of Part II, namely Proposition 5.1, provides the existence of such a lattice point for a fixed q with finitely many (only a few) exceptional values for w. For these exceptions, we perform a direct search over all possible lattice points to get τC (not lying inside R_w) and construct the counterexample as described in section 16. Note that a possible compensate change (section 15) can be found by a lattice search as well.

This construction of the actual counterexample and lattice search for the exceptions extends Algorithm 10.1; thus completes the proof of Proposition 2.5.

Applying this algorithm, the existence of counterexamples for all $w \geq 4$ is shown; the only exceptions are $q = 4$, $p \in \{-1, 1\}$ and $w = 6$ which are handled separately by Proposition 17.1.

Non-minimality of the cases $w = 2$ and $w = 3$ (and arbitrary $q \geq 3$) is proven in section 13. Minimality for $q = 2$ and $w \in \{2, 3\}$ is shown in [2, 3, 8]. This finishes the proof of Theorem 2.4. □

References

[1] R. M. Avanzi, *A note on the signed sliding window integer recoding and a left-to-right analogue*, Selected areas in cryptography, Lecture Notes in Comput. Sci., vol. 3357, Springer, Berlin, 2005, pp. 130–143, DOI 10.1007/978-3-540-30564-4_9. MR2180673

[2] R. M. Avanzi, C. Heuberger, and H. Prodinger, *Minimality of the Hamming weight of the τ-NAF for Koblitz curves and improved combination with point halving*, Selected areas in cryptography, Lecture Notes in Comput. Sci., vol. 3897, Springer, Berlin, 2006, pp. 332–344, DOI 10.1007/11693383_23. MR2241647

[3] R. M. Avanzi, C. Heuberger, and H. Prodinger, *Scalar multiplication on Koblitz curves using the Frobenius endomorphism and its combination with point halving: extensions and mathematical analysis*, Algorithmica **46** (2006), no. 3-4, 249–270, DOI 10.1007/s00453-006-0105-9. MR2291956

[4] R. Avanzi, C. Heuberger, and H. Prodinger, *Redundant τ-adic expansions I: non-adjacent digit sets and their applications to scalar multiplication*, Des. Codes Cryptogr. **58** (2011), no. 2, 173–202, DOI 10.1007/s10623-010-9396-6. MR2770310

[5] A. Baker and H. Davenport, *The equations $3x^2 - 2 = y^2$ and $8x^2 - 7 = z^2$*, Quart. J. Math. Oxford Ser. (2) **20** (1969), 129–137. MR0248079

[6] I. F. Blake, G. Seroussi, and N. P. Smart, *Elliptic Curves in Cryptography*, London Mathematical Society Lecture Note Series, vol. 265, Cambridge University Press, Cambridge, 2000. MR1771549

[7] J. W. S. Cassels, *An Introduction to the Geometry of Numbers*, Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete, Bd. 99 Springer-Verlag, Berlin-Göttingen-Heidelberg, 1959. MR0157947

[8] D. M. Gordon, *A survey of fast exponentiation methods*, J. Algorithms **27** (1998), no. 1, 129–146, DOI 10.1006/jagm.1997.0913. MR1613189

[9] C. Heuberger, *Redundant τ-adic expansions. II. Non-optimality and chaotic behaviour*, Math. Comput. Sci. **3** (2010), no. 2, 141–157, DOI 10.1007/s11786-009-0014-9. MR2608292

[10] C. Heuberger and D. Krenn, *Analysis of width-w non-adjacent forms to imaginary quadratic bases*, J. Number Theory **133** (2013), no. 5, 1752–1808, DOI 10.1016/j.jnt.2012.08.029. MR3007130

[11] C. Heuberger and D. Krenn, *Existence and optimality of w-non-adjacent forms with an algebraic integer base*, Acta Math. Hungar. **140** (2013), no. 1-2, 90–104. MR3123865

[12] C. Heuberger and D. Krenn, *Optimality of the width-w non-adjacent form: general characterisation and the case of imaginary quadratic bases* (English, with English and French summaries), J. Théor. Nombres Bordeaux **25** (2013), no. 2, 353–386. MR3228312

[13] J. Jedwab and C. J. Mitchell, *Minimum weight modified signed-digit representations and fast exponentiation*, Electron. Lett. **25** (1989), 1171–1172.

[14] N. Koblitz, *CM-curves with good cryptographic properties*, Advances in Cryptology—CRYPTO '91 (Santa Barbara, CA, 1991), Lecture Notes in Comput. Sci., vol. 576, Springer, Berlin, 1992, pp. 279–287, DOI 10.1007/3-540-46766-1_22. MR1243654

[15] N. Koblitz, *An elliptic curve implementation of the finite field digital signature algorithm*, Advances in Cryptology—CRYPTO '98 (Santa Barbara, CA, 1998), Lecture Notes in Comput. Sci., vol. 1462, Springer, Berlin, 1998, pp. 327–337, DOI 10.1007/BFb0055739. MR1670960

[16] E. M. Matveev, *An explicit lower bound for a homogeneous rational linear form in logarithms of algebraic numbers. II*, Izv. Ross. Akad. Nauk Ser. Mat. **64** (2000), no. 6, 125–180, DOI 10.1070/IM2000v064n06ABEH000314; English transl., Izv. Math. **64** (2000), no. 6, 1217–1269. MR1817252

[17] W. Meier and O. Staffelbach, *Efficient multiplication on certain nonsupersingular elliptic curves*, Advances in Cryptology—CRYPTO '92 (Santa Barbara, CA, 1992), Lecture Notes in Comput. Sci., vol. 740, Springer, Berlin, 1993, pp. 333–344, DOI 10.1007/3-540-48071-4_24. MR1287863

[18] A. Miyaji, T. Ono, and H. Cohen, *Efficient elliptic curve exponentiation*, Information and Communications Security. 1st International Conference, ICICS '97, Beijing, China, November 11–14, 1997. Proceedings (Yongfei Han, Tatsuaki Okamoto, and Sihan Qing, eds.), Lecture Notes in Comput. Sci., vol. 1334, Springer-Verlag, 1997, pp. 282–290.

[19] J. A. Muir and D. R. Stinson, *Minimality and other properties of the width-w nonadjacent form*, Math. Comp. **75** (2006), no. 253, 369–384, DOI 10.1090/S0025-5718-05-01769-2. MR2176404

[20] A. Pethő and B. M. M. de Weger, *Products of prime powers in binary recurrence sequences. I. The hyperbolic case, with an application to the generalized Ramanujan-Nagell equation*, Math. Comp. **47** (1986), no. 176, 713–727, DOI 10.2307/2008185. MR856715

[21] B. Phillips and N. Burgess, *Minimal weight digit set conversions*, IEEE Trans. Comput. **53** (2004), 666–677.

[22] G. W. Reitwiesner, *Binary arithmetic*, Advances in Computers, Vol. 1, Academic Press, New York, 1960, pp. 231–308. MR0122018

[23] P. Ribenboim, *Classical Theory of Algebraic Numbers*, Universitext, Springer-Verlag, New York, 2001. MR1821363

[24] The SageMath Developers, *SageMath Mathematics Software (Version 7.0)*, 2016, `http://www.sagemath.org`.

[25] J. A. Solinas, *An improved algorithm for arithmetic on a family of elliptic curves*, Advances in Cryptology — CRYPTO '97. 17th Annual International Cryptology Conference, Santa Barbara, CA, August 17–21, 1997. Proceedings (B. S. Kaliski, ed.), Lecture Notes in Comput. Sci., vol. 1294, Springer, Berlin, 1997, pp. 357–371.

[26] J. A. Solinas, *Efficient arithmetic on Koblitz curves*, Des. Codes Cryptogr. **19** (2000), no. 2-3, 195–249, DOI 10.1023/A:1008306223194. Towards a quarter-century of public key cryptography. MR1759617

Institut für Mathematik, Alpen-Adria-Universität Klagenfurt, Universitätsstrasse 65–67, 9020 Klagenfurt am Wörthersee, Austria
E-mail address: `math@danielkrenn.at`
E-mail address: `daniel.krenn@aau.at`

Fachbereich für Mathematik, University of Salzburg, Hellbrunnerstrasse 34, A-5020 Salzburg, Austria
E-mail address: `volker.ziegler@sbg.ac.at`