Moscow Journal of Combinatorics and Number Theory

2019
vol. 8 no. 4

msp

Moscow Journal of Combinatorics and Number Theory

msp.org/moscow

Cover design: Blake Knoll, Alex Scorpan and Silvio Levy

See inside back cover or msp.org/moscow for submission instructions.

The subscription price for 2019 is US $310/year for the electronic version, and $365/year (+$20, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Moscow Journal of Combinatorics and Number Theory (ISSN 2640-7361 electronic, 2220-5438 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

MJCNT peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY
mathematical sciences publishers
nonprofit scientific publishing
http://msp.org/

Moscow Journal of Combinatorics and Number Theory
Vol. 8, No. 4, 2019
dx.doi.org/10.2140/moscow.2019.8.289

Paramodular forms of level 16 and supercuspidal representations

Cris Poor, Ralf Schmidt and David S. Yuen

This work bridges the abstract representation theory of GSp(4) with recent computational techniques. We construct four examples of paramodular newforms whose associated automorphic representations have local representations at $p = 2$ that are supercuspidal. We classify all relevant irreducible, admissible, supercuspidal representations of GSp(4, $\mathbb{Q}_2$), and show that our examples occur at the lowest possible paramodular level, 16. The required theoretical and computational techniques include paramodular newform theory, Jacobi restriction, bootstrapping and Borcherds products.

The authors thank ICERM for the June 12–16, 2017, Collaborate@ICERM grant where a substantial part of this work was completed. We thank the Academic Computing Environment at Fordham University for the use of its servers.
MSC2010: 11F46, 11F70.
Keywords: Siegel modular forms, paramodular forms.

Introduction

This paper consists of a local part and a global part. In the local part we classify irreducible, admissible, supercuspidal representations of $\mathrm{GSp}(4, \mathbb{Q}_2)$ with trivial central character and small conductor. In particular, we prove that there exists a unique such supercuspidal $\mathrm{sc}(16)$ with (the exponent of the) conductor $a(\mathrm{sc}(16)) = 4$. In the global part we construct Siegel paramodular cusp forms of weights 9, 11, 13, and 14 and paramodular level 16 generating an automorphic representation with $\mathrm{sc}(16)$ as its 2-component. To the best of our knowledge, these are the first examples of Siegel paramodular forms generating automorphic representations with a supercuspidal component. Other types of local representations can be seen in [PSY 2018].

We give two approaches to the construction of $\mathrm{sc}(16)$. The first approach relies on the local Langlands correspondence for the groups $\mathrm{GL}(2)$, $\mathrm{GL}(4)$ and $\mathrm{GSp}(4)$. We first construct, via automorphic induction, a set of six supercuspidals of $\mathrm{GL}(2, E)$, where $E = \mathbb{Q}_2(\sqrt{5})$ is the unramified quadratic extension of $\mathbb{Q}_2$. Up to unramified twists, these are precisely the depth zero supercuspidals of $\mathrm{GL}(2, E)$. We automorphically induce again to obtain three supercuspidals of $\mathrm{GL}(4, \mathbb{Q}_2)$. These are precisely the three depth zero supercuspidals of $\mathrm{GL}(4, \mathbb{Q}_2)$ with trivial central character. Of these three, exactly one is a transfer from a representation of $\mathrm{GSp}(4, \mathbb{Q}_2)$. This representation of $\mathrm{GSp}(4, \mathbb{Q}_2)$ is the unique generic supercuspidal $\mathrm{sc}(16)$ with trivial central character and conductor 4. As a corollary to our construction, we obtain a complete list of all supercuspidals of $\mathrm{GSp}(4, \mathbb{Q}_2)$ with trivial central character and conductor ≤ 4; see Table 2. We also determine, via direct calculation, that the value of the ε-factor at $1/2$ of $\mathrm{sc}(16)$ is -1. This sign is important to know for global applications, as it will help us to identify $\mathrm{sc}(16)$ within the automorphic representations generated by paramodular forms.

Our second approach to $\mathrm{sc}(16)$ is via compact induction. The Langlands parameter $\mathrm{sc}(16)$, known from the first construction, is of the kind considered in [DeBacker and Reeder 2009]. The results of this paper then exhibit $\mathrm{sc}(16)$ as being compactly induced from a cuspidal representation κ_0 of $\mathrm{GSp}(4, \mathbb{Z}_2/2\mathbb{Z}_2)$ (inflated to $\mathrm{GSp}(4, \mathbb{Z}_2)$ and extended trivially to include the center). Since $\mathrm{GSp}(4, \mathbb{Z}/2\mathbb{Z}) \cong S_6$, the irreducible characters of this group are in bijection with the partitions of 6. The representation κ_0 corresponds to $(2, 2, 1, 1)$ and has dimension 9. It is the unique cuspidal, generic character of $\mathrm{GSp}(4, \mathbb{Z}/2\mathbb{Z})$.

We describe the passage from global paramodular forms to local supercuspidal representations. The automorphic representations studied here are generated by the adelic function canonically associated to a paramodular eigenform $f \in S_k(K(N))^{\mathrm{new}}$. The interesting local representations are classified by computing the Hecke eigenvalues of f at primes dividing the level N. In order to rigorously compute these eigenvalues, we span the Fricke eigenspace containing f, $S_k(K(N))^{\epsilon}$. Accurate upper bounds for the dimension of $S_k(K(N))^{\epsilon}$ are provided by *Jacobi restriction*, which classifies all possible Fourier–Jacobi coefficients from $S_k(K(N))^{\epsilon}$ to some sufficient order. Lower bounds are created by the technique of *bootstrapping*. Bootstrapping seeds the target space with a Borcherds product, and then generates a subspace that contains the seed and is stable under a good Hecke operator. Bootstrapping is run modulo an auxiliary prime, and the subtle point is that it does not directly compute the action of a good Hecke operator $T(q)$ on $S_k(K(N))^{\epsilon}$, but rather of a formal Hecke operator $\mathcal{T}(q)$ on the Jacobi restriction space of initial Fourier–Jacobi expansions.

Even with the relevant spaces spanned, the eigenvalues at the bad primes resist direct computation because they involve Fourier coefficients from more than one 1-dimensional cusp. As in [PSY 2018], this is overcome using the technique of *restriction to a modular curve*. We found symmetric f with a supercuspidal local component early on, but only found the antisymmetric example in $S_{14}(K(16))^-$ as the computations were becoming prohibitive.

1. Notation

For any commutative ring R, let

$$\mathrm{GSp}(4,R)=\{g\in\mathrm{GL}(4,R) : {}^tgJg=\mu(g)J,\ \text{for some } \mu(g)\in R^\times\},\qquad J=\begin{bmatrix} & 1_2\\ -1_2 & \end{bmatrix}.$$

The kernel of the multiplier homomorphism $\mu : \mathrm{GSp}(4,R)\to R^\times$ is the group $\mathrm{Sp}(4,R)$. The $\mathbb{C}$-vector space of Siegel modular forms of weight $k\in\mathbb{Z}$ for a subgroup $\Gamma\subseteq\mathrm{GSp}(4,\mathbb{R})$ commensurable with $\mathrm{Sp}(4,\mathbb{Z})$ is denoted by $M_k(\Gamma)$, the subspace of cusp forms by $S_k(\Gamma)$.

2. Supercuspidal representations of GSp(4, $\mathbb{Q}_2$) of small conductor

Let F be a non-archimedean local field of characteristic zero. Let $\mathfrak{o}$ be its ring of integers, $\mathfrak{p}$ the maximal ideal of $\mathfrak{o}$, and q the cardinality of the residue class field $\mathfrak{o}/\mathfrak{p}$. When there is more than one field involved, we sometimes write $\mathfrak{o}_F$, $\mathfrak{p}_F$, and q_F for clarity.

Let W_F be the Weil group of F, and W_F' the Weil–Deligne group. We refer to [Rohrlich 1994] or [Gross and Reeder 2010] for basic facts about the Weil and Weil–Deligne groups and their representations. If $\phi : W_F'\to\mathrm{GL}(n,\mathbb{C})$ is a representation of W_F', then we define the (exponent of the) *conductor* $a(\phi)$ of ϕ as in §10 of [Rohrlich 1994]. If π is an irreducible, admissible representation of $\mathrm{GL}(n,F)$, then the conductor of π is defined as $a(\pi)=a(\phi)$, where $\phi : W_F'\to\mathrm{GL}(n,\mathbb{C})$ is the Weil–Deligne representation corresponding to π via the local Langlands correspondence.

2.1. ***Discrete series parameters for* GSp(4).** The local Langlands correspondence (LLC) for $\mathrm{GL}(n)$ states that there is a bijection between isomorphism classes of irreducible, admissible representations π of $\mathrm{GL}(n,F)$ and Langlands parameters, i.e., conjugacy classes of admissible homomorphisms $\phi : W_F'\to\mathrm{GL}(n,\mathbb{C})$. This bijection satisfies a number of desirable properties. For example, if π corresponds to ϕ, then the central character of π corresponds to $\det(\phi)$ under the LLC for $\mathrm{GL}(1)$ (which is essentially the reciprocity law of local class field theory). Another property is that π is an essentially discrete series representation if and only if the image of ϕ is not contained in a proper Levi subgroup; such ϕ are therefore called discrete series parameters. Moreover, supercuspidal π correspond to irreducible ϕ.

The local Langlands correspondence is also a theorem for $\mathrm{GSp}(4)$; see [Gan and Takeda 2011]. The Langlands parameters are now admissible homomorphisms $\phi : W_F'\to\mathrm{GSp}(4,\mathbb{C})$, taken up to conjugacy by elements of $\mathrm{GSp}(4,\mathbb{C})$. A new phenomenon is that to one ϕ there now corresponds either a single representation π, as in the $\mathrm{GL}(n)$ case, or a set of two representations $\{\pi_1,\pi_2\}$. In either case we speak of the L-packet corresponding to ϕ. The size of the L-packet corresponding to ϕ equals the cardinality of S_ϕ/S_ϕ^0Z, where S_ϕ is the centralizer of the image of ϕ, S_ϕ^0 is its identity component, and Z is the center of $\mathrm{GSp}(4,\mathbb{C})$.

The LLC for $\mathrm{GSp}(4)$ is such that the central character of the representations in the L-packet of ϕ corresponds to the multiplier $\mu\circ\phi$. As in the $\mathrm{GL}(n)$ case, the L-packet corresponding to ϕ consists of essentially discrete series representations if and only if the image of ϕ is not contained in a proper Levi subgroup of $\mathrm{GSp}(4,\mathbb{C})$. It is also true that irreducible $\phi : W'_F \to \mathrm{GSp}(4,\mathbb{C})$ correspond to singleton supercuspidal L-packets. However, there are plenty of supercuspidals whose Langlands parameter is not irreducible.

To better understand L-parameters for supercuspidals, we recall some of the discussion of Section 7 of [Gan and Takeda 2011]. Let $\phi : W'_F \to \mathrm{GSp}(4,\mathbb{C})$ be a discrete series parameter for $\mathrm{GSp}(4)$, meaning the image of ϕ is not contained in a proper Levi subgroup of $\mathrm{GSp}(4,\mathbb{C})$. Such parameters are of one of two types (A) or (B).

Type (A): Viewed as a four-dimensional representation of W'_F, the map ϕ decomposes as $\phi_1 \oplus \phi_2$, where ϕ_1 and ϕ_2 are inequivalent indecomposable two-dimensional representations of W'_F with $\det(\phi_1) = \det(\phi_2)$. Explicitly, if $\phi_i(w) = \left[\begin{smallmatrix} a_i(w) & b_i(w) \\ c_i(w) & d_i(w) \end{smallmatrix}\right]$, then

$$\phi(w) = \begin{bmatrix} a_1(w) & & b_1(w) & \\ & a_2(w) & & b_2(w) \\ c_1(w) & & d_1(w) & \\ & c_2(w) & & d_2(w) \end{bmatrix}.$$

In this case the packet associated to ϕ consists of two elements, a generic representation π^{gen} and a non-generic π^{ng}. The common central character of these two representations corresponds to $\det(\phi_1) = \det(\phi_2)$. There are three subcases:

- **($\mathbf{A}_1$)**: Both ϕ_1 and ϕ_2 are irreducible. In this case π^{gen} and π^{ng} are both supercuspidal.
- **($\mathbf{A}_2$)**: One of ϕ_1, ϕ_2 is irreducible, and the other is reducible (but indecomposable). In this case π^{gen} is a representation of type XIa in the classification of [Roberts and Schmidt 2007]; it sits inside a representation induced from a supercuspidal representation of the Levi component of the Siegel parabolic subgroup. The non-generic π^{ng} is supercuspidal; it is a representation of type XIa* in the notation of [Roberts and Schmidt 2016].
- **($\mathbf{A}_3$)**: Both ϕ_1 and ϕ_2 are reducible (but indecomposable). In this case π^{gen} is a representation of type Va in the classification of [Roberts and Schmidt 2007]; it sits inside a representation induced from the Borel subgroup. The non-generic π^{ng} is supercuspidal; it is a representation of type Va* in the notation of [Roberts and Schmidt 2016].

Hence π^{ng} is always supercuspidal, but π^{gen} is only supercuspidal for class (A_1). Note that, by Theorem 3.4.3 of [Roberts and Schmidt 2007], non-generic supercuspidals do not contain paramodular vectors of any level. Hence, supercuspidals of the form π^{ng} cannot occur as local components in automorphic representations attached to paramodular cusp forms.

Type (B): Viewed as a four-dimensional representation of W'_F, the map ϕ is indecomposable. In this case there is a single representation π attached to ϕ, and this π is generic. Via the inclusion $\mathrm{GSp}(4,\mathbb{C}) \hookrightarrow \mathrm{GL}(4,\mathbb{C})$ we may view ϕ as the Langlands parameter of a discrete series representation Π of $\mathrm{GL}(4,F)$. By the definitions involved, Π is the image of π under the functorial lifting from $\mathrm{GSp}(4)$ to $\mathrm{GL}(4)$ coming from the embedding $\mathrm{GSp}(4,\mathbb{C}) \hookrightarrow \mathrm{GL}(4,\mathbb{C})$ of dual groups. Again there are three subcases:

- **($\mathbf{B}_1$)**: ϕ is irreducible as a four-dimensional representation. In this case π is supercuspidal.
- **($\mathbf{B}_2$)**: $\phi = \varphi \otimes \mathrm{sp}(2)$ with an irreducible two-dimensional representation φ of W_F, and $\mathrm{sp}(2)$ being the special indecomposable two-dimensional representation of W'_F. In this case π is a representation of

type IXa; see Section 2.4 of [Roberts and Schmidt 2007]. This π sits inside a representation induced from a supercuspidal representation of the Levi component of the Klingen parabolic subgroup.

- **(B$_3$)**: $\phi = \xi \otimes \mathrm{sp}(4)$ with a one-dimensional representation ξ of W_F. Then π is a twist of the Steinberg representation $\mathrm{St}_{\mathrm{GSp}(4)}$ (type IVa in the classification of [loc. cit.]).

Hence π is supercuspidal only for class (B$_1$), i.e., if ϕ is irreducible. In this case π transfers to a supercuspidal representation Π of $\mathrm{GL}(4, F)$.

2.2. *Counting supercuspidals for* GL(2) *and* GL(4). We see from the parameters exhibited in the previous section that, in order to understand supercuspidal representations of $\mathrm{GSp}(4, F)$, we need to understand supercuspidal representations of $\mathrm{GL}(2, F)$ and $\mathrm{GL}(4, F)$, or equivalently, two-dimensional and four-dimensional irreducible representations of W_F. In this section we count the number of supercuspidals of $\mathrm{GL}(2, F)$ and $\mathrm{GL}(4, F)$ with small conductor.

The *conductor* $a(\pi)$ of an irreducible, admissible representation of $\mathrm{GL}(n, F)$ is by definition the Artin conductor $a(\phi)$ of its Langlands parameter ϕ; see §10 of [Rohrlich 1994]. Here, we always mean the *exponent* of the conductor, so that $a(\pi) = a(\phi)$ is a non-negative integer. Another measure of complexity is the *depth* $d(\pi)$, as defined in [Moy and Prasad 1994; 1996]. For supercuspidals, there is an easy relationship between depth and conductor, given by

$$d(\pi) = \frac{a(\pi) - n}{n}; \tag{1}$$

see Proposition 2.2 of [Lansky and Raghuram 2003]. The set of supercuspidals of a fixed conductor is invariant under unramified twisting.

The smallest conductor that can occur for a supercuspidal representation of $\mathrm{GL}(n, F)$ is $a(\pi) = n$. By (1), these are the depth zero supercuspidals. If π is one such supercuspidal, and χ is an unramified character, then the twist $\chi\pi$ is also a depth zero supercuspidal. For a positive integer n, let Z_n be the (finite) set of isomorphism classes of depth zero supercuspidals of $\mathrm{GL}(n, F)$ up to unramified twists. It is known that Z_n is in bijection with the set of $\mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q)$ orbits of length n in the group of characters of $\mathbb{F}_{q^n}^\times$; see Section 8 of [Deligne and Lusztig 1976] and Section 6 of [Moy and Prasad 1996]. It is an exercise to show that

$$\#Z_2 = \tfrac{1}{2}q(q-1), \qquad \#Z_4 = \tfrac{1}{4}q^2(q^2-1). \tag{2}$$

Note that if $q = 2$, then every tamely ramified character of $F^\times$ is unramified. Hence, in this case, every element of Z_n is represented by a unique depth zero supercuspidal with trivial central character. The reason is that depth zero supercuspidals are compactly induced from representations of ZK, where the representation on $K = \mathrm{GL}(2, \mathfrak{o})$ is inflated from a cuspidal representation of $\mathrm{GL}(2, \mathfrak{o}/\mathfrak{p})$. If $\mathfrak{o}/\mathfrak{p}$ has only two elements, then every representation of K thus obtained has trivial central character. In particular, we see from (2) that $\mathrm{GL}(2, \mathbb{Q}_2)$ has exactly one depth zero supercuspidal with trivial central character, and $\mathrm{GL}(4, \mathbb{Q}_2)$ has exactly three depth zero supercuspidals with trivial central character.

For a unitary character ω of $F^\times$, let S_ω be the set of isomorphism classes of depth zero supercuspidals of $\mathrm{GL}(2, F)$ with central character ω. By Proposition 3.4 of [Tunnell 1978], $\#S_\omega = 0$ if $a(\omega) \geq 2$. If

$a(\omega) \le 1$, then, by (4-1) of [Knightly and Ragsdale 2014],

$$\#S_\omega = \begin{cases} \frac{1}{2}(q-1) & \text{if } q \text{ is odd and } \omega^{(q-1)/2} \text{ is trivial,} \\ \frac{1}{2}(q+1) & \text{if } q \text{ is odd and } \omega^{(q-1)/2} \text{ is nontrivial,} \\ \frac{1}{2}q & \text{if } q \text{ is even.} \end{cases} \tag{3}$$

2.3. ***Depth zero supercuspidals of*** $\mathbf{GL(2,\ \mathbb{Q}_2(\sqrt{5}))}$**.** In this section let $E = \mathbb{Q}_2(\sqrt{5})$ be the unramified quadratic extension of $\mathbb{Q}_2$, and let L be the unramified quadratic extension of E. Note that 2 is a uniformizer both in E and in L. Let $\mathbb{F}_{p^n}$ be the field with p^n elements. The residue class field of E is $\mathbb{F}_4$, and the residue class field of L is $\mathbb{F}_{16}$. The polynomial $X^4 + X + 1 \in \mathbb{F}_2[X]$ is irreducible, so that

$$\mathbb{F}_{16} \cong \mathbb{F}_2[X]/(X^4 + X + 1).$$

Let $\bar{y}$ be the image of X via this isomorphism. Then $\mathbb{F}_{16} = \mathbb{F}_2(\bar{y})$, and $\bar{y}$ satisfies $\bar{y}^4 = \bar{y} + 1$. Clearly, the order of $\bar{y}$ in $\mathbb{F}_{16}^\times$ is not 3 or 5, so that $\bar{y}$ is a generator of the cyclic group $\mathbb{F}_{16}^\times$. The element $\bar{y}^5$ is then a generator of the cyclic group $\mathbb{F}_4^\times$. Let y be an element of $\mathfrak{o}_L^\times$ mapping to $\bar{y}$ under the projection $\mathfrak{o}_L^\times \to \mathbb{F}_{16}^\times$.

Let $\bar{\eta}$ be the character of $\mathbb{F}_{16}^\times$ determined by $\bar{\eta}(\bar{y}) = e^{2\pi i/15}$. For $r \in \mathbb{Z}/15\mathbb{Z}$ we define a character η_r of $L^\times$ by lifting $\bar{\eta}^r$ to $\mathfrak{o}_L^\times$ and setting $\eta_r(2) = -1$.

Let θ be the generator of $\mathrm{Gal}(L/\mathbb{Q}_2)$ that induces the map $x \mapsto x^2$ on $\mathbb{F}_{16}$. Then θ^2 generates $\mathrm{Gal}(L/E)$. We have $\eta_r^\theta = \eta_{2r}$. (If $\sigma \in \mathrm{Gal}(L/F)$ and π is a representation of $\mathrm{GL}(n, F)$, then π^σ is the representation of $\mathrm{GL}(n, F)$ defined by $\pi^\sigma(g) = \pi(\sigma(g))$.)

Consider automorphic induction $AI = AI_{L/E}$; see [Henniart and Herb 1995]. Recall that AI takes characters ξ of $L^\times$ to irreducible, admissible representations ρ of $\mathrm{GL}(2, E)$. By Proposition 4.5 of the same reference, the central character of ρ is given by $\chi_{L/E}(\xi|_{E^\times})$, where $\chi_{L/E}$ is the quadratic character of $E^\times$ corresponding to the extension L/E. On the Galois side, AI corresponds to induction of parameters, i.e., the parameter of ρ is

$$\phi_\rho = \mathrm{ind}_{W_L}^{W_E}(\xi).$$

This parameter is irreducible, i.e., ρ is supercuspidal, if and only if ξ is not $\mathrm{Gal}(L/E)$-invariant. We have $a(\rho) = 2a(\xi)$ by the conductor formula (a2) in §10 of [Rohrlich 1994].

We now consider $AI_{L/E}(\eta_r)$ for $r \in \{1, \dots, 15\}$. This representation is supercuspidal if and only if $\eta_{4r} \ne \eta_r$, which translates into $5 \nmid r$. Since $a(\eta) = 1$, we have $a(AI_{L/E}(\eta_r)) = 2$ for $r \ne 0$. The central character ω of $AI_{L/E}(\eta_r)$ is determined by $\omega(2) = 1$ and $\omega(y^5) = \eta_{5r}(y) = e^{2\pi i r/3}$. Hence, if we let ω_j be the character of $E^\times$ which is trivial on $1 + \mathfrak{p}_E$ and satisfies $\omega_j(2) = 1$ and $\omega_j(y^5) = e^{2\pi i(j-1)/3}$, then $\omega_1, \omega_2, \omega_3$ are the possible central characters of the $AI_{L/E}(\eta_r)$. We have $\omega_1^\theta = \omega_1$ and $\omega_2^\theta = \omega_3$. Considering Langlands parameters, it is easy to see that the $\mathrm{Gal}(E/\mathbb{Q}_2)$-conjugate of $AI(\xi)$ is given by $AI(\xi)^\theta = AI(\xi^\theta)$, and the contragredient is $AI(\xi)^\vee = AI(\xi^{-1})$.

Table 1 lists the supercuspidal representations of the form $AI(\eta_r)$. For each possible central character ω_j, there are two supercuspidals, which we denote by ρ_{ja} and ρ_{jb}. Note from (2) that there are exactly six depth zero supercuspidals of $\mathrm{GL}(2, E)$ up to unramified twists. The following lemma implies that the six representations $\{\rho_{1a}, \rho_{1b}, \rho_{2a}, \rho_{2b}, \rho_{3a}, \rho_{3b}\}$ represent these six classes of depth zero supercuspidals up to unramified twists. Note that having exactly two depth zero supercuspidals for a given central character ω_j is consistent with (3).

ξ	$AI(\xi)$	ω	$AI(\xi)^\theta$	$AI(\xi)^\vee$
η_3 or η_{12}	ρ_{1a}	ω_1	ρ_{1b}	ρ_{1a}
η_6 or η_9	ρ_{1b}	ω_1	ρ_{1a}	ρ_{1b}
η or η_4	ρ_{2a}	ω_2	ρ_{3a}	ρ_{3b}
η_7 or η_{13}	ρ_{2b}	ω_2	ρ_{3b}	ρ_{3a}
η_2 or η_8	ρ_{3a}	ω_3	ρ_{2a}	ρ_{2b}
η_{11} or η_{14}	ρ_{3b}	ω_3	ρ_{2b}	ρ_{2a}

Table 1. Representatives for the depth zero supercuspidals of $\mathrm{GL}(2, E)$ up to unramified twists. The first column shows $\mathrm{Gal}(L/E)$-orbits of length 2 of the characters $\xi = \eta_r$. The ω column shows the central character of the representation $AI_{L/E}(\xi)$. The columns $AI(\xi)^\theta$ and $AI(\xi)^\vee$ show the $\mathrm{Gal}(E/\mathbb{Q}_2)$-conjugate and contragredient of $AI_{L/E}(\xi)$, respectively.

Lemma 2.3.1. *Let $j \in \{1, 2, 3\}$.*

i) *The representation ρ_{ja} is not a twist of ρ_{jb}.*

ii) *Let $\rho = \rho_{ja}$ or $\rho = \rho_{jb}$. Then ρ^θ is not isomorphic to a twist of $\rho^\vee$.*

iii) *Let $\rho, \rho' \in \{\rho_{1a}, \rho_{1b}, \rho_{2a}, \rho_{2b}, \rho_{3a}, \rho_{3b}\}$. Then ρ is not an unramified twist of ρ', unless $\rho = \rho'$.*

Proof. i) Assume that $\rho_{ja} = \chi \otimes \rho_{jb}$ for some character χ of $E^\times$; we will obtain a contradiction. Taking central characters on both sides, we see that $\chi^2 = 1$. We have $a(\chi) \le 1$ by Proposition 3.4 of [Tunnell 1978].

Assume that $a(\chi) = 0$. Then χ is either the trivial character, or $\chi = \chi_{L/E}$, the unique nontrivial, unramified, quadratic character of $E^\times$. In either case $\chi \otimes \rho_{jb} = \rho_{jb}$, a contradiction.

Assume that $a(\chi) = 1$. Then χ induces a nontrivial character of $\mathfrak{o}_E^\times/(1 + \mathfrak{p}_E)$. In particular, the image of $\chi|_{\mathfrak{o}_E^\times}$ consists of the third roots of unity, contradicting $\chi^2 = 1$.

ii) follows from i) and Table 1.

iii) Assume that ρ is an unramified twist of ρ'. Then the restrictions of the central characters of ρ and ρ' to $\mathfrak{o}_E^\times$ coincide. Hence $\rho = \rho_{j*}$ and $\rho' = \rho_{j*}$ with the same j. By i), we conclude $\rho = \rho'$. □

Lemma 2.3.2. *Let L be the unramified extension of degree 4 over $\mathbb{Q}_2$. Let the characters η_r of $L^\times$ be defined as above. Then, for $\xi \in \{\eta_3, \eta_6, \eta_9, \eta_{12}\}$,*

$$\varepsilon(1/2, \xi, \psi_L) = -1. \tag{4}$$

Here, $\psi_L = \psi \circ \mathrm{tr}_{L/\mathbb{Q}_2}$, where ψ is a character of $\mathbb{Q}_2$ that is trivial on $\mathbb{Z}_2$ but not on $2^{-1}\mathbb{Z}_2$.

Proof. Let

$$\mathbb{F}_{16} \cong \mathbb{F}_2[X]/(X^4 + X + 1),$$

and let $\bar{y}$ be the element corresponding to X, as at the beginning of this section. The Frobenius of the extension $\mathbb{F}_{16}/\mathbb{F}_2$ is given by squaring, so that

$$\mathrm{tr}_{\mathbb{F}_{16}/\mathbb{F}_2}(x) = x + x^2 + x^4 + x^8$$

for any $x \in \mathbb{F}_{16}$. Using this formula and $\bar{y}^4 = \bar{y} + 1$, it is easy to calculate the trace of any element of $\mathbb{F}_{16}$. The results are as follows:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\mathrm{tr}_{\mathbb{F}_{16}/\mathbb{F}_2}(\bar{y}^i)$	0	0	1	0	0	1	1	0	1	0	1	1	1	1	0

(5)

Let $\xi \in \{\eta_3, \eta_6, \eta_9, \eta_{12}\}$. By the formula ($\epsilon 3$) in §11 of [Rohrlich 1994],

$$\varepsilon(1/2, \xi, \psi_L) = q_L^{-a(\xi)/2} \int_{\varpi_L^{-a(\xi)} \mathfrak{o}_L^\times} \xi^{-1}(x) \psi_L(x)\, dx. \tag{6}$$

For this formula to hold, it is important that ψ_L has conductor $\mathfrak{o}_L$, which is the case for our additive character. The element ϖ_L is a uniformizer; in our case we may take $\varpi_L = 2$. We further have $a(\xi) = 1$ and $q_L = 16$, so that

$$\begin{aligned}\varepsilon(1/2, \xi, \psi_L) &= \frac{1}{4} \int_{2^{-1}\mathfrak{o}_L^\times} \xi^{-1}(x)\psi_L(x)\, dx = \frac{1}{4}\, |2^{-1}|_L \int_{\mathfrak{o}_L^\times} \xi^{-1}(2^{-1}x)\psi_L(2^{-1}x)\, dx \\ &= -4 \int_{\mathfrak{o}_L^\times} \xi^{-1}(x)\psi_L(2^{-1}x)\, dx = -4 \operatorname{vol}(1+\mathfrak{p}_L) \sum_{x \in \mathfrak{o}_L^\times/(1+\mathfrak{p}_L)} \xi^{-1}(x)\psi_L(2^{-1}x) \\ &= -\frac{1}{4} \sum_{x \in \mathfrak{o}_L^\times/(1+\mathfrak{p}_L)} \xi^{-1}(x)\psi(2^{-1}\mathrm{tr}_{L/\mathbb{Q}_2}(x)).\end{aligned}$$

We have

$$\psi(2^{-1}\mathrm{tr}_{L/\mathbb{Q}_2}(x)) = \begin{cases} 1 & \text{if } \mathrm{tr}_{L/\mathbb{Q}_2}(x) \in 2\mathbb{Z}_2 \quad (\text{equivalently, if } \mathrm{tr}_{\mathbb{F}_{16}/\mathbb{F}_2}(\bar{x}) = 0), \\ -1 & \text{if } \mathrm{tr}_{L/\mathbb{Q}_2}(x) \in \mathbb{Z}_2^\times \quad (\text{equivalently, if } \mathrm{tr}_{\mathbb{F}_{16}/\mathbb{F}_2}(\bar{x}) = 1). \end{cases}$$

Hence, using (5),

$$\begin{aligned}\varepsilon(1/2, \xi, \psi_L) &= -\frac{1}{4}\Bigg(\sum_{i \in \{1,2,4,5,8,10,15\}} \xi^{-1}(y^i) - \sum_{i \in \{3,6,7,9,11,12,13,14\}} \xi^{-1}(y^i) \Bigg) \\ &= -\tfrac{1}{4}\big(\zeta + \zeta^2 + \zeta^4 + \zeta^5 + \zeta^8 + \zeta^{10} + \zeta^{15} - \zeta^3 - \zeta^6 - \zeta^7 - \zeta^9 - \zeta^{11} - \zeta^{12} - \zeta^{13} - \zeta^{14}\big),\end{aligned}$$

where $\zeta = \xi^{-1}(y)$, a primitive fifth root of unity. Using $\zeta^5 = 1$ and $1 + \zeta + \zeta^2 + \zeta^3 + \zeta^4 = 0$, this simplifies to

$$\begin{aligned}\varepsilon(1/2, \xi, \psi_L) &= -\tfrac{1}{4}\big(\zeta + \zeta^2 + \zeta^4 + 1 + \zeta^3 + 1 + 1 - \zeta^3 - \zeta - \zeta^2 - \zeta^4 - \zeta - \zeta^2 - \zeta^3 - \zeta^4\big) \\ &= -\tfrac{1}{4}\big(3 - \zeta - \zeta^2 - \zeta^3 - \zeta^4\big) = -1.\end{aligned}$$

This concludes the proof. □

2.4. ***Supercuspidals of* GSp(4, $\mathbb{Q}_2$) *with small conductor.*** As in the previous section, let E be the unramified quadratic extension of $\mathbb{Q}_2$. Let θ be the nontrivial element of $\mathrm{Gal}(E/\mathbb{Q}_2)$. We now consider automorphic induction $AI = AI_{E/\mathbb{Q}_2}$. Recall that AI takes irreducible, admissible representations ρ of $\mathrm{GL}(2, E)$ to irreducible, admissible representations π of $\mathrm{GL}(4, \mathbb{Q}_2)$. By Proposition 4.5 of [Henniart and Herb 1995], the central characters ω_ρ and ω_π are related by $\omega_\pi = \omega_\rho|_{\mathbb{Q}_2^\times}$. If ϕ_ρ is the parameter of ρ, then the parameter of π is

$$\phi_\pi = \mathrm{ind}_{W_E}^{W_{\mathbb{Q}_2}}(\phi_\rho).$$

Assume that ρ is supercuspidal, or equivalently, that ϕ_ρ is irreducible. Then π is supercuspidal if and only if $\rho \neq \rho^\theta$, where the Galois conjugate ρ^θ is defined by $\rho^\theta(g) = \rho(g^\theta)$ for $g \in \mathrm{GL}(2, E)$. In other words, ϕ_π is irreducible if and only if $\phi_\rho \neq \phi_\rho^\theta$, where $\phi_\rho^\theta(w) = \phi_\rho(\theta w \theta^{-1})$ for $w \in W_E$ (here we think of θ as an element of $W_{\mathbb{Q}_2}$ that is not in W_E). Also, we have $AI(\rho) = AI(\rho^\theta)$.

We apply $AI = AI_{E/\mathbb{Q}_2}$ to the supercuspidal representations of GL(2, E) listed in Table 1. It follows from this table that

$$AI(\rho_{1a}) = AI(\rho_{1b}), \qquad AI(\rho_{2a}) = AI(\rho_{3a}), \qquad AI(\rho_{2b}) = AI(\rho_{3b}), \tag{7}$$

and these are supercuspidal representations of GL(4, $\mathbb{Q}_2$). They all have trivial central character. By the conductor formula for induced representations of the Weil group, see (a2) in §10 of [Rohrlich 1994], they have conductor 4. It follows that the representations in (7) are precisely the three depth zero supercuspidals of GL(4, $\mathbb{Q}_2$) with trivial central character; see Section 2.2.

We will next determine which of the three supercuspidals in (7) are transfers from GSp(4). For any p-adic field F, an irreducible, admissible representation π of GL(4, F) is a transfer from GSp(4, F) if and only if its parameter $\phi_\pi : W'_F \to \mathrm{GL}(4, \mathbb{C})$, after suitable conjugation, has image in GSp(4, $\mathbb{C}$). Assume this is the case, and consider the exterior square map $\bigwedge^2 : \mathrm{GL}(4, \mathbb{C}) \to \mathrm{GL}(6, \mathbb{C})$. Since the composition of $\bigwedge^2$ with the inclusion $\mathrm{GSp}(4, \mathbb{C}) \hookrightarrow \mathrm{GL}(4, \mathbb{C})$ decomposes as the direct sum of a five-dimensional and a one-dimensional representation of GSp(4, $\mathbb{C}$), it follows that $\bigwedge^2 \circ \phi_\pi$ contains a one-dimensional representation of W'_F.

The following lemma was spelled out in a preprint version of [Gan and Takeda 2011] but not in the published version. We include a proof here.

Lemma 2.4.1. *Let E/F be a quadratic extension of p-adic fields. Let θ be an element of W_F that is not in W_E. Let (ϕ, V) be an irreducible two-dimensional representation of W_E, and let $\phi^\theta(w) = \phi(\theta w \theta^{-1})$ for $w \in W_E$. Then*

$$\textstyle\bigwedge^2 \left(\mathrm{ind}_{W_E}^{W_F}(\phi)\right) = U \oplus \mathrm{ind}_{W_E}^{W_F}(\det(\phi)),$$

where U is a 4-dimensional representation of W_F whose restriction to W_E is isomorphic to $\phi \otimes \phi^\theta$.

Proof. As a model for $\phi := \mathrm{ind}_{W_E}^{W_F}(\phi)$, we may take $V \oplus V$, with action

$$\phi(w)(v_1 \oplus v_2) = \phi(w)v_1 \oplus \phi^\theta(w)v_2 \quad (w \in W_E), \qquad \phi(\theta)(v_1 \oplus v_2) = v_2 \oplus \phi(\theta^2)v_1. \tag{8}$$

If spaces V_1 and V_2 carry an action of a group G, then

$$\textstyle\bigwedge^2 (V_1 \oplus V_2) \cong \bigwedge^2 V_1 \,\oplus\, (V_1 \otimes V_2) \,\oplus\, \bigwedge^2 V_2$$

as G-spaces. It follows that, as a W_E-representation,

$$\textstyle\bigwedge^2 \left(\mathrm{ind}_{W_E}^{W_F}(\phi)\right) = \det(\phi) \oplus (\phi \otimes \phi^\theta) \oplus \det(\phi)^\theta,$$

It is easy to see that $\det(\phi)\oplus\det(\phi)^\theta$ is invariant under the action of θ, and that in fact this two-dimensional space is isomorphic to $\mathrm{ind}_{W_E}^{W_F}(\det(\phi))$ as a W_F-representation. The space U realizing $\phi\otimes\phi^\theta$ is also invariant under θ. □

Lemma 2.4.2. *The representations* $AI_{E/\mathbb{Q}_2}(\rho_{2a})$ *and* $AI_{E/\mathbb{Q}_2}(\rho_{2b})$ *appearing in* (7) *are not transfers from* $\mathrm{GSp}(4,\mathbb{Q}_2)$.

Proof. Let $\rho=\rho_{2a}$ or ρ_{2b}. Let $\phi : W_E\to\mathrm{GL}(2,\mathbb{C})$ be the parameter of ρ. Then the parameter of $AI_{E/\mathbb{Q}_2}(\rho)$ is $\mathrm{ind}_{W_E}^{W_{\mathbb{Q}_2}}(\phi)$. By Lemma 2.4.1,

$$\bigwedge^2\big(\mathrm{ind}_{W_E}^{W_{\mathbb{Q}_2}}(\phi)\big)=U\oplus\mathrm{ind}_{W_E}^{W_{\mathbb{Q}_2}}(\det(\phi)),$$

where U is isomorphic to $\phi\otimes\phi^\theta$ as a W_E-representation. By Lemma 2.3.1 ii), the space U is irreducible, even as a W_E-representation. Since $\det(\phi)=\omega_2$ is not $\mathrm{Gal}(E/\mathbb{Q}_2)$-invariant, the two-dimensional $\mathrm{ind}_{W_E}^{W_{\mathbb{Q}_2}}(\det(\phi))$ is irreducible as a $W_{\mathbb{Q}_2}$-representation. Hence $\bigwedge^2\big(\mathrm{ind}_{W_E}^{W_{\mathbb{Q}_2}}(\phi)\big)$ does not contain any one-dimensional component. By our remarks above, $AI_{E/\mathbb{Q}_2}(\rho)$ cannot be a transfer from $\mathrm{GSp}(4,\mathbb{Q}_2)$. □

Theorem 2.4.3. *The group* $\mathrm{GSp}(4,\mathbb{Q}_2)$ *admits a unique generic supercuspidal representation* $\mathsf{sc}(16)$ *with conductor* $a(\mathsf{sc}(16))=4$ *and trivial central character. As a four-dimensional representation of* $W_{\mathbb{Q}_2}$*, the Langlands parameter of* $\mathsf{sc}(16)$ *is*

$$\phi_{\mathsf{sc}(16)}=\mathrm{ind}_{W_L}^{W_{\mathbb{Q}_2}}(\xi),\tag{9}$$

where L is the unramified extension of $\mathbb{Q}_2$ *of degree* 4, *and* ξ *is any character of* $L^\times$ *with the following properties*: ξ *is trivial on* $1+\mathfrak{p}_L$; *the values of the restriction of* ξ *to* $\mathfrak{o}_L^\times$ *are the fifth roots of unity*; $\xi(2)=-1$. *We have* $\varepsilon(1/2,\mathsf{sc}(16),\psi)=-1$, *where* ψ *is a character of* $\mathbb{Q}_2$ *which is trivial on* $\mathbb{Z}_2$ *but not on* $2^{-1}\mathbb{Z}_2$.

Proof. Let π be a generic supercuspidal representation of $\mathrm{GSp}(4,\mathbb{Q}_2)$ with $a(\pi)=4$ and trivial central character. The requirement that π be generic excludes supercuspidals of type Va* and XIa*; these are the ones with parameters of type (A_2) and (A_3), as defined in Section 2.1. Assume that π has a parameter of type (A_1); we will obtain a contradiction. Parameters of type (A_1) are of the form $\phi_1\oplus\phi_2$, where ϕ_1,ϕ_2 are inequivalent irreducible, two-dimensional representations of $W_{\mathbb{Q}_2}$ with $\det(\phi_1)=\det(\phi_2)=1$. Since $a(\pi)=4$, we must have $a(\phi_1)=a(\phi_2)=2$. Hence ϕ_1 and ϕ_2 correspond to supercuspidals of $\mathrm{GL}(2,\mathbb{Q}_2)$ with conductor 2 and trivial central character. By (3), there exists only one such supercuspidal. Hence $\phi_1\cong\phi_2$, a contradiction.

By our considerations in Section 2.1, the parameter of π is of type (B_1), i.e., irreducible as a four-dimensional representation. Hence π transfers to a supercuspidal representation π' on $\mathrm{GL}(4,\mathbb{Q}_2)$ with trivial central character and $a(\pi')=4$. It follows that π' is one of the representations in (7). By Lemma 2.4.2 we must have $\pi'=AI_{E/\mathbb{Q}_2}(\rho_{1a})=AI_{E/\mathbb{Q}_2}(\rho_{1b})$, where E is the unramified quadratic extension of $\mathbb{Q}_2$. This shows that, as a four-dimensional representation, the parameter of π is

$$\mathrm{ind}_{W_E}^{W_{\mathbb{Q}_2}}(\phi_a)=\mathrm{ind}_{W_E}^{W_{\mathbb{Q}_2}}(\phi_b),\tag{10}$$

where ϕ_* is the parameter of ρ_{1*}. By the considerations on p. 284/285 of [Roberts 2001], there exists a

$a(\pi)$	π	type	generic	$\varepsilon(1/2,\pi)$	$L(s,\pi)^{-1}$
2	$\delta^*([\xi,\nu\xi],\nu^{-1/2})$	Va*	no	-1	$1-q^{-2s-1}$
3	$\delta^*(\nu^{1/2}\tau_2,\nu^{-1/2})$	XIa*	no	1	$1-q^{-s-1/2}$
	$\delta^*(\nu^{1/2}\tau_2,\nu^{-1/2}\xi)$	XIa*	no	-1	$1+q^{-s-1/2}$
4	$\delta^*(\nu^{1/2}\tau_3,\nu^{-1/2})$	XIa*	no	-1	$1-q^{-s-1/2}$
	$\delta^*(\nu^{1/2}\tau_3,\nu^{-1/2}\xi)$	XIa*	no	-1	$1+q^{-s-1/2}$
	$\delta^*(\nu^{1/2}\xi\tau_3,\nu^{-1/2})$	XIa*	no	1	$1-q^{-s-1/2}$
	$\delta^*(\nu^{1/2}\xi\tau_3,\nu^{-1/2}\xi)$	XIa*	no	1	$1+q^{-s-1/2}$
	sc(16)		yes	-1	1

Table 2. The supercuspidals π of GSp(4, $\mathbb{Q}_2$) with conductor $a(\pi)\le 4$ and trivial central character. The character ξ is the unique nontrivial, unramified, quadratic character of $\mathbb{Q}_2^\times$. The representation τ_2 is the unique supercuspidal of GL(2, $\mathbb{Q}_2$) with trivial central character and conductor 2. The representation τ_3 (resp. $\xi\tau_3$) is the unique supercuspidal of GL(2, $\mathbb{Q}_2$) with trivial central character, conductor 3 and root number 1 (resp. -1). The representation sc(16) is the one from Theorem 2.4.3. The non-generic supercuspidals $\delta^*(\dots)$ share an L-packet with the generic square-integrable representations $\delta(\dots)$ of type Va resp. XIa; see Section 4.6 of [Roberts and Schmidt 2016].

unique symplectic structure on the space of $\mathrm{ind}_{W_E}^{W_{\mathbb{Q}_2}}(\phi_*)$ for which $W_{\mathbb{Q}_2}$ acts with trivial similitude. We proved that the parameter of π is uniquely determined. The uniqueness and existence of π now follows from the local Langlands correspondence for GSp(4, $\mathbb{Q}_2$).

Let η_j be as in Table 1. Then $\eta_3,\eta_6,\eta_9,\eta_{12}$ are precisely the characters ξ of $L^\times$ with $\xi(2)=-1$, trivial on $1+\mathfrak{p}_L$, and such that the values of the restriction of ξ to $\mathfrak{o}_L^\times$ are the fifth roots of unity. Inducing η_3 or η_{12} to W_E gives the parameter ϕ_a of ρ_{1a}, and inducing η_6 or η_9 to W_E gives the parameter ϕ_b of ρ_{1b}. Hence (9) follows by transitivity of induction.

We have $\varepsilon(1/2,\pi,\psi)=\varepsilon(1/2,\xi,\psi_L)$ by Corollary 4 to Theorem 5.6 of [Henniart and Herb 1995], or by (ϵ2) in §11 of [Rohrlich 1994]. Hence the assertion about $\varepsilon(1/2,\pi,\psi)$ follows from Lemma 2.3.2. □

Corollary 2.4.4. *Table 2 contains a complete list of all the irreducible, admissible, supercuspidal representations π of* GSp(4, $\mathbb{Q}_2$) *with trivial central character and conductor $a(\pi)\le 4$.*

Proof. Let π be an irreducible, admissible, supercuspidal representations of GSp(4, $\mathbb{Q}_2$) with trivial central character and conductor $a(\pi)\le 4$. Assume first that π is generic. Then π cannot be of type Va* or XIa*. Equivalently, the Langlands parameter ϕ of π cannot be of type (A$_2$) or (A$_3$). Assume that ϕ is of type (A$_1$), so that $\phi=\phi_1\oplus\phi_2$ with irreducible, two-dimensional, inequivalent representations ϕ_1,ϕ_2 of $W_{\mathbb{Q}_2}$ for which $\det(\phi_1)=\det(\phi_2)=1$. Since $a(\phi_1), a(\phi_2)\ge 2$ and $a(\phi)\le 4$, we have $a(\phi_1)=a(\phi_2)=2$ and $a(\phi)=4$. It follows that π must be the representation sc(16) of Theorem 2.4.3. But then π transfers to a supercuspidal on GL(4, $\mathbb{Q}_2$), contradicting the reducibility of ϕ. This contradiction shows that ϕ cannot be of type (A$_1$). Alternatively, one can argue that, by (3), there is only one supercuspidal τ_2 of GL(2, $\mathbb{Q}_2$) with conductor 2 and trivial central character, contradicting the inequivalence of ϕ_1 and ϕ_2.

We proved that a generic supercuspidal π of $\mathrm{GSp}(4,\mathbb{Q}_2)$ with trivial central character and conductor $a(\pi)\le 4$ must have a parameter ϕ of type (B_1). Hence π transfers to a supercuspidal on $\mathrm{GL}(4,\mathbb{Q}_2)$ and must have $a(\pi)=4$. Thus π is the representation $\mathrm{sc}(16)$ of Theorem 2.4.3.

Next assume that π is a non-generic supercuspidal of $\mathrm{GSp}(4,\mathbb{Q}_2)$ with trivial central character and conductor $a(\pi)\le 4$. Then π must have a parameter ϕ of type (A). Since $a(\phi)\le 4$, the argument above shows that ϕ cannot be of type (A_1), so that ϕ is of type (A_2) or (A_3).

Assume that ϕ is of type (A_3). By definition, $\phi=\phi_1\oplus\phi_2$, where ϕ_1,ϕ_2 are reducible but indecomposable, inequivalent, and satisfy $\det(\phi_1)=\det(\phi_2)=1$. Hence ϕ_i is the parameter of $\sigma_i\mathrm{St}_{\mathrm{GL}(2)}$ for distinct quadratic characters σ_1,σ_2 of $\mathbb{Q}_2^\times$. The restrictions on the conductors imply that σ_1 and σ_2 must both be unramified; see the proposition in §10 of [Rohrlich 1994]. Hence one of σ_1,σ_2 is trivial, and the other is the unique nontrivial, unramified, quadratic character ξ of $\mathbb{Q}_2^\times$. (This ξ is given by the local Hilbert symbol $(\cdot,5)$.) The corresponding π is the representation $\delta^*([\xi,\nu\xi],\nu^{-1/2})$ of type Va^*.

Assume that ϕ is of type (A_2). By definition, $\phi=\phi_1\oplus\phi_2$, where ϕ_1 is irreducible and ϕ_2 is the parameter of $\sigma\mathrm{St}_{\mathrm{GL}(2)}$ for some character σ of $\mathbb{Q}_2^\times$. Moreover $\det(\phi_1)=\det(\phi_2)=1$. Since $a(\phi)\le 4$, the character σ must be unramified, so that either $\sigma=1$ or $\sigma=\xi$. In both cases $a(\phi_2)=1$, which implies $a(\phi_1)\in\{2,3\}$. There is only one possible ϕ_1 with $a(\phi_1)=2$, namely the parameter of τ_2, the unique supercuspidal of $\mathrm{GL}(2,\mathbb{Q}_2)$ with trivial central character and conductor 2; see (3). From this ϕ_1 we therefore obtain two supercuspidals π with $a(\pi)=3$. Using the notation of [Roberts and Schmidt 2016], these are the representations $\delta^*(\nu^{1/2}\tau_2,\nu^{-1/2})$ and $\delta^*(\nu^{1/2}\tau_2,\xi\nu^{-1/2})$ of type XIa^*.

Finally, consider the case $a(\phi_1)=3$. By Theorem 3.9 of [Tunnell 1978], there are exactly two possibilities for ϕ_1. One corresponds to a supercuspidal representation τ_3 of $\mathrm{GL}(2,\mathbb{Q}_2)$ with trivial central character, $a(\tau_3)=3$ and $\varepsilon(1/2,\tau_3)=1$. The other corresponds to the twist $\xi\tau_3$, which is distinguished from τ_3 by the value of the ε-factor $\varepsilon(1/2,\xi\tau_3)=-1$. The two possibilities of ϕ_1, together with the two possibilities for σ, lead to four supercuspidals π of type XIa^*.

For the non-generic representations, the values of the L- and ϵ-factors in Table 2 can be read off Tables A.8 and A.9 of [Roberts and Schmidt 2007]. Note that Va^* has the same factors as Va, since they constitute a two-element L-packet; similarly for XIa and XIa^*. The ε-factor for $\mathrm{sc}(16)$ is given in Theorem 2.4.3. The L-factor for $\mathrm{sc}(16)$ is 1, since the parameter of $\mathrm{sc}(16)$ is irreducible. □

We refer to Section 4 of [Roberts and Schmidt 2016] for a construction of the representations of type Va^* and XIa^* in terms of the theta correspondence. Note that the representation of type Va^* occurring in Table 2 is invariant under twisting by the unramified character ξ.

2.5. ***The representation* sc(16) *via compact induction.*** We give an alternative construction of the supercuspidal representation $\mathrm{sc}(16)$ by employing compact induction. Consider the Langlands parameter $\phi_{\mathrm{sc}(16)}$ of $\mathrm{sc}(16)$ given in (9). After choosing a suitable basis of $\mathrm{ind}_{W_L}^{W_{\mathbb{Q}_2}}(\xi)$ we may think of $\phi_{\mathrm{sc}(16)}$ as a map $W_{\mathbb{Q}_2}\to\mathrm{GSp}(4,\mathbb{C})$. The image lies in fact in $\mathrm{Sp}(4,\mathbb{C})$, the dual group of $G=\mathrm{SO}(5)\cong\mathrm{PGSp}(4)$, so that, if we wish, we may work in a semisimple context.

In this section we consider the Vogan L-packet of $\phi_{\mathrm{sc}(16)}$. Recall that a Vogan L-packet may contain representations across all *pure inner forms* of a group; see [Vogan 1993] or the overview in Section 3 of [Gross and Prasad 1992]. As explained in Section 8 of [Gross and Reeder 2010], the split group $\mathrm{SO}(2n+1)$ has a unique non-split pure inner form $\mathrm{SO}^*(2n+1)$. We will see that the L-packet of $\phi_{\mathrm{sc}(16)}$

has two elements, one being a representation of $\mathrm{SO}(5, \mathbb{Q}_2) \cong \mathrm{PGSp}(4, \mathbb{Q}_2)$ (this is our sc(16)), the other one a representation of $\mathrm{SO}^*(5, \mathbb{Q}_2)$.

The parameter $\phi_{\mathsf{sc}(16)} : W_{\mathbb{Q}_2} \to \mathrm{Sp}(4, \mathbb{C})$ is *discrete* in the sense that its image has finite centralizer. It is *tame* in the sense that the image of wild inertia is trivial; this is because the character $\xi : L^\times \to \mathbb{C}^\times$ is trivial on $1 + \mathfrak{p}_L$. Moreover, $\phi_{\mathsf{sc}(16)}$ is in *general position*, meaning the image of tame inertia is generated by a regular, semisimple element. Hence $\phi_{\mathsf{sc}(16)}$ is among the Langlands parameters considered in [DeBacker and Reeder 2009]. The construction in [DeBacker and Reeder 2009] attaches a Vogan L-packet to each tame, discrete Langlands parameter in general position. In the context of GSp(4), the paper [Lust 2013] assures that the packets thus obtained coincide with the L-packets defined in [Gan and Takeda 2011] and [Gan and Tantono 2014].

The centralizer C_ϕ of the image of $\phi_{\mathsf{sc}(16)} : W_{\mathbb{Q}_2} \to \mathrm{Sp}(4, \mathbb{C})$ is precisely the center $\pm I_4$ of $\mathrm{Sp}(4, \mathbb{C})$. The work [DeBacker and Reeder 2009] attaches to each irreducible character ρ of C_ϕ a depth-zero supercuspidal representation on a pure inner form of the group under consideration. In our case, going through the definitions shows that the trivial character of C_ϕ gives rise to a representation of $\mathrm{SO}(5, \mathbb{Q}_2)$, and the nontrivial character to a representation of $\mathrm{SO}^*(5, \mathbb{Q}_2)$. We will concentrate on the former, since (by [Lust 2013]) this is our supercuspidal sc(16).

As explained in Section 4.4 of [DeBacker and Reeder 2009], each irreducible character ρ of C_ϕ gives rise to an orbit of vertices in the Bruhat–Tits building of $G = \mathrm{PGSp}(4)$ over $\mathbb{Q}_2$. By Lemma 6.2.1 of [DeBacker and Reeder 2009], these vertices are hyperspecial if and only if ρ is trivial. It is exactly the hyperspecial vertices that lead to *generic* depth-zero supercuspidals, consistent with the fact that sc(16) is generic.

We may work with the hyperspecial vertex x_0 whose associated parahoric subgroup is $p(K)$, where $K = \mathrm{GSp}(4, \mathbb{Z}_2)$ and $p : \mathrm{GSp}(4, \mathbb{Q}_2) \to G(\mathbb{Q}_2)$ is the projection. Let G_0 be the reductive group over the residue class field $\mathfrak{f} = \mathbb{F}_2$ attached to x_0, so that $\mathsf{G}_0(\mathfrak{f}) \cong p(K)/p(K)^+$, where $p(K)^+$ is the pro-unipotent radical of $p(K)$. In our case $p(K)^+$ is a principal congruence subgroup, and $\mathsf{G}_0 = \mathrm{Sp}(4)$. The construction of sc(16) is then as follows. The parameter $\phi_{\mathsf{sc}(16)}$ determines an $\mathfrak{f}$-minisotropic maximal torus T_0 in G_0. The restriction of $\phi_{\mathsf{sc}(16)}$ to tame inertia defines a character θ of $\mathsf{T}_0(\mathfrak{f})$ via the tame local Langlands correspondence for tori. Since $\phi_{\mathsf{sc}(16)}$ is in general position, the character θ will be in general position in the sense of Definition 5.15 of [Deligne and Lusztig 1976]. Deligne–Lusztig induction therefore yields an irreducible, cuspidal character

$$\kappa_0 = \pm R_{T,\theta} \tag{11}$$

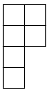
of $\mathsf{G}_0(\mathfrak{f}) \cong \mathrm{Sp}(4, \mathfrak{f})$. Let κ be the inflation of κ_0 to $p(K)$ via $\mathsf{G}_0(\mathfrak{f}) \cong p(K)/p(K)^+$. Then

$$\mathrm{sc}(16) = \text{c-Ind}_{p(K)}^{G(\mathbb{Q}_2)}(\kappa), \tag{12}$$

where we identify representations of $G(\mathbb{Q}_2)$ with representations of $\mathrm{GSp}(4, \mathbb{Q}_2)$ with trivial central character. Alternatively, we can first pull back κ to a character of K, extend it trivially to ZK, where Z is the center of $\mathrm{GSp}(4, \mathbb{Q}_2)$, and compactly induce to $\mathrm{GSp}(4, \mathbb{Q}_2)$. By Proposition 6.6 of [Moy and Prasad 1996], the induced representation in (12) is irreducible and supercuspidal.

Making things explicit, one finds that T_0 is the maximal torus corresponding to the conjugacy class consisting of length 2 elements in the 8-element Weyl group of G_0; see Section 3.3 of [Carter 1985] for the correspondence between conjugacy classes in the Weyl group and maximal tori. The group $\mathsf{T}_0(\mathfrak{f})$ is cyclic of order 5. The characters θ of $\mathsf{T}_0(\mathfrak{f})$ in general position are precisely the isomorphisms of this group with the fifth roots of unity. By Corollary 7.2 of [Deligne and Lusztig 1976], the character κ_0 in (11) has degree 9.

It is an exercise in elementary character theory to show that $\mathrm{Sp}(4,\mathfrak{f})$ has exactly one irreducible, cuspidal representation κ_0 of dimension 9, and that this representation is generic; see [Enomoto 1972] for information on the characters of $\mathrm{Sp}(4,\mathbb{F}_{2^n})$. This κ_0 corresponds to the irreducible character with Young diagram



(13)

under the isomorphism of $\mathrm{Sp}(4,\mathfrak{f})$ with the symmetric group S_6 described in Section 3.5.2 of [Wilson 2009]. There is in fact only one other irreducible, cuspidal character of $\mathrm{Sp}(4,\mathfrak{f})$, namely the one-dimensional sign character under the isomorphism $\mathrm{Sp}(4,\mathfrak{f})\cong S_6$.

To summarize, $\mathrm{sc}(16)$ is a depth-zero supercuspidal representation of $\mathrm{GSp}(4,\mathbb{Q}_2)$ which may be constructed as follows. Take the unique irreducible, cuspidal character κ_0 of $\mathrm{Sp}(4,\mathfrak{f})$ that is not one-dimensional; it has dimension 9 and is generic. Inflate κ_0 to a representation κ of $K=\mathrm{GSp}(4,\mathbb{Z}_2)$ and extend it to ZK by making it trivial on the center Z of $\mathrm{GSp}(4,\mathbb{Q}_2)$. Then we have $\mathrm{sc}(16)=\text{c-Ind}_{ZK}^{\mathrm{GSp}(4,\mathbb{Q}_2)}(\kappa)$. The Vogan L-packet of $\mathrm{sc}(16)$ contains an additional representation which lives on the non-split inner form of $\mathrm{GSp}(4)$.

3. Paramodular cusp forms of weight $k\le 14$ and level $N=16$

A good reference for the notation in this section and hereafter is [PSY 2018]. For each $N\in\mathbb{N}$, the paramodular group, $K(N)$, and its normalizing Fricke involution, μ_N, are defined by

$$K(N)=\begin{bmatrix} * & N* & * & * \\ * & * & * & */N \\ * & N* & * & * \\ N* & N* & N* & * \end{bmatrix}\cap\mathrm{Sp}(4,\mathbb{Q}),\ *\in\mathbb{Z}\,;\qquad \mu_N=\frac{1}{\sqrt{N}}\begin{bmatrix} 0 & -N & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & N & 0 \end{bmatrix}.$$

Let $S_k(K(N))^{\epsilon}$ for $\epsilon=\pm$ denote the Fricke eigenspace of $S_k(K(N))$ with eigenvalue ± 1, so that we have the decomposition $S_k(K(N))=S_k(K(N))^{+}\oplus S_k(K(N))^{-}$. In the case where N is a power of a prime, the Fricke sign is also the Atkin–Lehner sign at that prime. The Gritsenko lift is an injective linear map from $J_{k,N}^{\mathrm{cusp}}$ to $S_k(K(N))^{\epsilon}$ for $\epsilon=(-1)^k$. Paramodular forms that are not Gritsenko lifts will be called *nonlifts*.

We are searching for a supercuspidal paramodular form, i.e., a newform $f\in S_k(K(N))$ whose associated adelic representation has a supercuspidal local component. Since non-generic supercuspidals do not admit non-zero paramodular vectors by Theorem 3.4.3 of [Roberts and Schmidt 2007], a supercuspidal coming from a paramodular newform f is necessarily generic. In particular, f must be a nonlift. By Table 2, among 2-powers, the smallest N for which f can be supercuspidal is $N=16$. By Corollary 7.5.5 of [Roberts and Schmidt 2007], the value of the ε-factor at $1/2$ of an irreducible, admissible, generic representation coincides with the eigenvalue of the Atkin–Lehner involution on the newform. It therefore follows from Table 2 that if $S_k(K(16))$ contains a supercuspidal form, it must occur in $S_k(K(16))^{-}$. Hence, we pay special attention to these spaces.

Our first goal is to find all the nonlift newforms in $S_k(K(16))^{\pm}$ for $k \le 14$. In order to separate the nonlift newforms from the nonlift oldforms, we also find all the nonlift eigenforms in $S_k(K(N))$ for $k \le 14$ and $N \in \{1, 2, 4, 8\}$; we separate these eigenforms into their Fricke eigenspaces as well. The dimensions of $S_k(K(N))$ are known for $N \in \{1, 2, 4\}$, see [Igusa 1962; Ibukiyama and Onodera 1997; Poor and Yuen 2013]. Comparing with the known [Skoruppa and Zagier 1989] dimensions of Jacobi cusp forms $J_{k,N}^{\mathrm{cusp}}$, we see that $S_k(K(N))$ for $N \in \{1, 2\}$ and $k \le 14$ does not have any nonlifts. Thus we need only consider $N \in \{4, 8, 16\}$ in this section. Our first task is to compute the dimension of each of these spaces, and this will entail finding upper and lower bounds that are equal.

3.1. ***Paramodular forms and Fourier expansions.*** A paramodular form $f \in S_k(K(N))$ has a Fourier expansion

$$f(\Omega) = \sum_t a(t; f) e\left(\langle \Omega, t\rangle\right)$$

where the sum is over $t \in \mathcal{X}_2(N) = \left\{\left[\begin{smallmatrix} n & r/2 \\ r/2 & Nm \end{smallmatrix}\right] > 0 : n, r, m \in \mathbb{Z}\right\}$ and where $\langle \Omega, t\rangle = \mathrm{tr}(\Omega t)$. The similarity group $\left\{u \in \mathrm{GL}(2, \mathbb{R}) : \left[\begin{smallmatrix} u & 0 \\ 0 & u^* \end{smallmatrix}\right] \in K(N)\right\}$ equals $\hat{\Gamma}^0(N) = \left\langle \Gamma^0(N), \left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right]\right\rangle$, where, as usual, $\Gamma^0(N) = \left\{\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathrm{SL}(2, \mathbb{Z}) : b \equiv 0 \bmod N\right\}$, and hence the Fourier coefficients satisfy the following relations amongst themselves: for $t[u] = {}^t u t u$,

$$a(t[u]; f) = \det(u)^k a(t; f), \quad \text{for all} \quad u \in \hat{\Gamma}^0(N). \tag{14}$$

Another set of important relations among the Fourier coefficients comes from the Fricke involution μ_N; we have $a(t; f|\mu_N) = a(\mathrm{Twin}(t); f)$ for

$$t = \left[\begin{smallmatrix} n & r/2 \\ r/2 & Nm \end{smallmatrix}\right], \qquad \mathrm{Twin}(t) = \left[\begin{smallmatrix} m & -r/2 \\ -r/2 & Nn \end{smallmatrix}\right], \tag{15}$$

so that $t \mapsto \mathrm{Twin}(t)$ gives the action of μ_N on the Fourier coefficients. Therefore Fricke eigenforms obey the additional conditions

$$a\left(\mathrm{Twin}(t); f\right) = \epsilon\, a(t; f), \qquad \text{for } f \in S_k(K(N))^{\epsilon}. \tag{16}$$

Note that twinning stabilizes $\mathcal{X}_2(N)$ and respects $\hat{\Gamma}^0(N)$-classes. These observations follow from the equation $\mathrm{Twin}(t) = F_N t\, {}^t F_N$, for $F_N = \frac{1}{\sqrt{N}}\left[\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right]$, the elliptic Fricke involution on $\Gamma_0(N)$. We may view the Fourier expansion as a map $\mathrm{FE} : S_k(K(N)) \to \prod_{t \in \mathcal{X}_2(N)} \mathbb{C}$ that sends f to $(a(t; f))_{t \in \mathcal{X}_2(N)}$. Relations (14) and (16) above show that the image of $S_k(K(N))^{\epsilon}$ under FE lies in a very special subspace.

For a ring $R \subseteq \mathbb{C}$, we define $S_k(K(N))(R) \subseteq S_k(K(N))$ to be the R-module of paramodular cusp forms $f \in S_k(K(N))$ with $a(t; f) \in R$ for all $t \in \mathcal{X}_2(N)$. Fundamental results of Shimura [1975] show that general spaces of modular forms have integral bases, i.e., a basis with integral Fourier coefficients.

The natural reduction map $\mathrm{R_p} : \mathbb{Z} \to \mathbb{F}_p$ allows us to define modular forms over $\mathbb{F}_p$, a concept useful for both theory and computations: $S_k(K(N))(\mathbb{F}_p) = \mathrm{R_p} \circ \mathrm{FE}\left(S_k(K(N))(\mathbb{Z})\right)$. Thus paramodular forms over $\mathbb{F}_p$ are formal series with coefficients in $\mathbb{F}_p$ and the Fourier expansion map $\mathrm{FE} : S_k(K(N))(\mathbb{F}_p) \to \prod_{t \in \mathcal{X}_2(N)} \mathbb{F}_p$ is really the identity map. From the existence of an integral basis, it follows from the structure theorem for finitely generated $\mathbb{Z}$-modules that

$$\dim_{\mathbb{C}} S_k(K(N))^{\epsilon} = \mathrm{rank}_{\mathbb{Z}} S_k(K(N))^{\epsilon}(\mathbb{Z}) = \dim_{\mathbb{F}_p} S_k(K(N))^{\epsilon}(\mathbb{F}_p).$$

For odd primes p, we have the direct sum $S_k(K(N))(\mathbb{F}_p) = S_k(K(N))^{+}(\mathbb{F}_p) \oplus S_k(K(N))^{-}(\mathbb{F}_p)$.

3.2. ***Good Hecke operators and their action on Fourier coefficients.*** A Hecke operator is called *good* when its similitude is prime to the level. For each prime q not dividing N, we use the good Hecke operator $T(q) : S_k(K(N)) \to S_k(K(N))$ defined as follows. Decompose $K(N)\,\mathrm{diag}(1, 1, q, q)K(N) = \cup_j K(N)\gamma_j$ into a union of distinct cosets. For $f \in S_k(K(N))$, set $f|T(q) = \sum_j f|\gamma_j$, which is again in $S_k(K(N))$. Since $T(q)$ commutes with the Fricke involution μ_N, $T(q)$ also stabilizes $S_k(K(N))^\epsilon$. The action of $T(q)$ on the Fourier expansion of f is given by

$$a(t; f|T(q)) = a(qt; f) + q^{k-2} a\big(q^{-1}\, t\big[\begin{smallmatrix} q & 0 \\ 0 & 1 \end{smallmatrix}\big]; f\big) + q^{k-2} \sum_{j \bmod q} a\big(q^{-1}\, t\big[\begin{smallmatrix} 1 & 0 \\ j & q \end{smallmatrix}\big]; f\big) + q^{2k-3} a(q^{-1}t; f). \tag{17}$$

For $k \geq 2$, this equation shows that $T(q)$ stabilizes $S_k(K(N))^\epsilon(R)$ and is R-linear for subrings R of $\mathbb{C}$. On $S_k(K(N))^\epsilon(\mathbb{F}_p)$, the reduction of $T(q)$, $T(q)_p$, is defined by $\big(\mathrm{R_p} \circ \mathrm{FE}(f)\big)\,|T(q)_p = \mathrm{R_p} \circ \mathrm{FE}\,(f|T(q))$ and also obeys equation (17).

A possible source of confusion is that equation (17) is valid for the *classical* normalization of the slash, setting $\sigma = \left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right] \in \mathrm{GSp}(4, \mathbb{R})^+$ with similitude $\mu = \mu(\sigma) = \det(\sigma)^{1/2}$,

$$(f|_k\sigma)(\Omega) = \mu^{2k-3} \det(C\Omega + D)^{-k} f\left((A\Omega + B)(C\Omega + D)^{-1}\right).$$

In contrast, representation theory employs the *scalar invariant* slash where the power of the similitude is μ^k instead of μ^{2k-3}. The tension between these normalizations is real because local Euler factors depend only upon the local representation for the scalar invariant action of the Hecke algebra, whereas $T(q)$ is uniformly defined over $\mathbb{Z}$ for weights $k \geq 2$ only for the classical action. Our concession to this tension is to write the scalar invariant action of the left and the classical action on the right, so that $f|T(q) = q^{k-3}T(q)f$.

3.3. ***Fourier–Jacobi expansions, Jacobi forms, and Jacobi Hecke operators.*** The Fourier expansion of a paramodular cusp form $f \in S_k(K(N))$ may be rearranged to give the Fourier–Jacobi expansion, setting $\Omega = \left[\begin{smallmatrix} \tau & z \\ z & \omega \end{smallmatrix}\right] \in \mathcal{H}_2$, and $q = e(\tau)$, $\zeta = e(z)$,

$$f(\Omega) = \sum_{j=1}^{\infty} \phi_j(\tau, z) e(Nj\omega), \tag{18}$$

$$\phi_j(\tau, z) = \sum_{\substack{n, r \in \mathbb{Z} \\ 4nNj > r^2}} a\big(\big[\begin{smallmatrix} n & r/2 \\ r/2 & Nj \end{smallmatrix}\big]; f\big) q^n \zeta^r. \tag{19}$$

When we want to indicate the dependence of the ϕ_j on f we will write $\phi_j(\tau, z; f)$ instead of $\phi_j(\tau, z)$, or $\phi_j(f)$ instead of ϕ_j. We recall the definition of a Jacobi form and the following subgroups, for rings $R \subseteq \mathbb{C}$,

$$P_{2,1}(R) = \begin{bmatrix} * & 0 & * & * \\ * & * & * & * \\ * & 0 & * & * \\ 0 & 0 & 0 & * \end{bmatrix} \cap \mathrm{Sp}(4, R); \quad GP_{2,1}(R) = \begin{bmatrix} * & 0 & * & * \\ * & * & * & * \\ * & 0 & * & * \\ 0 & 0 & 0 & * \end{bmatrix} \cap \mathrm{GSp}(4, R).$$

A *Jacobi form* $\phi \in J_{k,m}$ of weight $k \in \mathbb{Z}$ and index $m \in \mathbb{Z}_{\geq 0}$ is a holomorphic function $\phi : \mathcal{H} \times \mathbb{C} \to \mathbb{C}$ such that the associated function $E_m\phi : \mathcal{H}_2 \to \mathbb{C}$ given by $(E_m\phi)(\Omega) = \phi(\tau, z)e(m\omega)$ is invariant under $P_{2,1}(\mathbb{Z})$, and is bounded on domains of the type $\{\Omega \in \mathcal{H}_2 : \mathrm{Im}\,\Omega > Y_o\}$. The boundedness condition is essential and, given the other assumptions, is equivalent to a Fourier expansion for ϕ of the form $\phi(\tau, z) = \sum_{n,r \in \mathbb{Z}:\, n \geq 0,\, 4nm \geq r^2} c(n, r; \phi) q^n \zeta^r$. For *Jacobi cusp forms* $\phi \in J_{k,m}^{\mathrm{cusp}}$, we require $4mn > r^2$. For a *weakly holomorphic* $\psi \in J_{k,m}^{\mathrm{wh}}$ we drop the boundedness condition and require $n \gg -\infty$. Indices with $4mn \le r^2$ are called *singular*. Spaces of Jacobi forms have integral bases by [Eichler and Zagier 1985] and so we may define $J_{k,m}^{\mathrm{cusp}}(R)$ for R a subring of $\mathbb{C}$ or for $\mathbb{F}_p$ as in the case of paramodular forms.

The subgroup $K_\infty(N) = P_{2,1}(\mathbb{Q}) \cap K(N)$ stabilizes the Fourier–Jacobi expansion (18) term by term, so that each $\phi_j \in J_{k,Nj}^{\mathrm{cusp}}$ is a Jacobi form and the Fourier coefficients of the ϕ_j are

$$c(n,r;\phi_j) = a\left(\left[\begin{smallmatrix} n & r/2 \\ r/2 & Nj \end{smallmatrix}\right]; f\right). \tag{20}$$

The Fourier–Jacobi expansion defines a map

$$\mathrm{FJ} : S_k(K(N)) \to \prod_{j=1}^{\infty} J_{k,Nj}^{\mathrm{cusp}}, \qquad f \mapsto \sum_{j=1}^{\infty} \phi_j\, \xi^{Nj}, \tag{21}$$

where we have let $\xi = e(\omega)$ and identified the sum on the right with the vector (ϕ_j).

The infinite direct product $\prod_{j=1}^{\infty} J_{k,Nj}^{\mathrm{cusp}}$ is an inverse limit with respect to the projection maps

$$\mathrm{proj}_d^u : \bigoplus_{j=1}^{u} J_{k,Nj}^{\mathrm{cusp}} \to \bigoplus_{j=1}^{d} J_{k,Nj}^{\mathrm{cusp}}, \qquad \text{for } d \le u.$$

We also define $\mathrm{proj}_d^\infty : \prod\limits_{j=1}^{\infty} J_{k,Nj}^{\mathrm{cusp}} \to \bigoplus\limits_{j=1}^{d} J_{k,Nj}^{\mathrm{cusp}}$. The projection onto the first u Fourier–Jacobi coefficients

$$\mathrm{proj}_u^\infty \circ \mathrm{FJ} : S_k(K(N))^\epsilon \to \bigoplus_{j=1}^{u} J_{k,Nj}^{\mathrm{cusp}} \tag{22}$$

injects for sufficiently large u and algorithms to find u_0 such that the map (22) injects for $u \ge u_0$ may be found in [Breeding et al. 2016]. When N is a prime power for example, u_0 is roughly $Nk/5$ and Table 3 displays u_0 for $1 \le k \le 14$ and $N \in \{4, 8, 16\}$. We write $S_k(K(N))^\epsilon[u]$ for the projection of $S_k(K(N))^\epsilon$ onto its first u Fourier–Jacobi coefficients, i.e.,

$$S_k(K(N))^\epsilon[u] = \mathrm{proj}_u^\infty \circ \mathrm{FJ}\left(S_k(K(N))^\epsilon\right).$$

One cannot take an arbitrary sequence of Jacobi forms ϕ_j and obtain the Fourier–Jacobi expansion $\sum_{j=1}^{\infty} \phi_j \xi^{Nj}$ of some paramodular form. Indeed, the Fourier–Jacobi coefficients of a paramodular Fricke eigenform satisfy the following symmetries. Let $f \in S_k(K(N))^\epsilon$ have the Fourier–Jacobi expansion $\sum_{j=1}^{\infty} \phi_j \xi^{Nj}$. Then

for all $t_1 = \left[\begin{smallmatrix} n_1 & r_1/2 \\ r_1/2 & Nm_1 \end{smallmatrix}\right]$, $t_2 = \left[\begin{smallmatrix} n_2 & r_2/2 \\ r_2/2 & Nm_2 \end{smallmatrix}\right] \in \mathcal{X}_2(N)$, and $u \in \hat{\Gamma}^0(N)$,

$$t_1[u] = t_2 \quad \Longrightarrow \quad c(n_1, r_1; \phi_{m_1}) = \det(u)^k c(n_2, r_2; \phi_{m_2}), \tag{23}$$

and

$$\text{for all } t = \left[\begin{smallmatrix} n & r/2 \\ r/2 & Nm \end{smallmatrix}\right] \in \mathcal{X}_2(N), \qquad c(n,r;\phi_m) = (-1)^k \epsilon\, c(m,r;\phi_n). \tag{24}$$

Equations (23) and (24) are consequences of (14) and (16). We refer to equation (24) as the *involution conditions*. Formal series of Jacobi forms that satisfy (23) and (24) and converge in an appropriate sense are in fact Fourier–Jacobi expansions of paramodular forms; see [Ibukiyama et al. 2013].

	u_0				u_1^+, u_1^-		
k	$K(4)$	$K(8)$	$K(16)$	k	$K(4)$	$K(8)$	$K(16)$
1	0	0	0	1	0	0	0
2	0	0	2	2	0	0	0
3	0	1	4	3	0	0	0
4	0	2	7	4	0	0	1, 0
5	1	3	9	5	0	0, 1	0, 1
6	1	4	11	6	0	1, 0	2, 0
7	2	5	14	7	0, 1	0, 1	0, 2
8	3	6	16	8	1, 0	1, 0	2, 0
9	4	8	18	9	0, 1	0, 2	1, 3
10	4	9	21	10	1, 0	2, 0	3, 1
11	5	10	23	11	0, 2	1, 2	2, 4
12	5	11	25	12	2, 0	3, 1	4, 2
13	6	12	28	13	0, 2	2, 3	3, 4
14	6	13	30	14	2, 0	3, 2	5, 3

Table 3. A sufficient number u_0 to make projection from $S_k(K(N))^\epsilon$ onto the first u_0 Jacobi coefficients injective. An improved number u_1^ϵ is given in the second set.

Following [Gritsenko 1995], we present the action of $T(q)$ on the Fourier–Jacobi coefficients of a paramodular cusp form in terms of the Jacobi raising and lowering operators, V_q and W_q. The raising operator $V_q : J_{k,m} \to J_{k,mq}$ is defined, for primes q, by

$$\left(\phi|V_q\right)(\tau,z) = q^{k-1}\phi(q\tau,qz) + \frac{1}{q}\sum_{\lambda \bmod q} \phi\left(\frac{\tau+\lambda}{q}, z\right),$$

or equivalently by

$$c(n,r;\phi|V_q) = q^{k-1}c(\frac{n}{q},\frac{r}{q};\phi) + c(qn,r;\phi), \tag{25}$$

as in [Eichler and Zagier 1985]. The lowering operators $W_q : J_{k,m} \to J_{k,\frac{m}{q}}$ were introduced in a special case in [Kohnen and Skoruppa 1989]. Their image is zero when the prime q does not divide m. When q divides m, we have

$$(\phi|W_q)(\tau,z) = q^{k-2}\sum_{\lambda \bmod q} \phi(q\tau, z+\lambda\tau)e\left(\frac{m}{q}(2\lambda z+\lambda^2\tau)\right) + q^{-2}\sum_{\lambda,\mu \bmod q}\phi\left(\frac{\tau+\lambda}{q},\frac{z+\mu}{q}\right),$$

or equivalently

$$c(n,r;\phi|W_q) = c(qn,qr;\phi) + q^{k-2}\sum_{\lambda \bmod q} c\left(\frac{n+\lambda r+\frac{m}{q}\lambda^2}{q}, \frac{r+2\frac{m}{q}\lambda}{q};\phi\right). \tag{26}$$

The invariance properties of the raising and lowering operators, i.e., that they send Jacobi forms to Jacobi forms, can be obtained by considering them as the Hecke operators $V_q = K_\infty(N)\,\mathrm{diag}(q, q, 1, 1)K_\infty(N)$ and $W_q = K_\infty(N)\,\mathrm{diag}(1, 1, q, q)K_\infty(N)$ for the noncommutative Jacobi Hecke algebra for $K_\infty(N)$ inside $GP_{2,1}(\mathbb{Q})$, see [Gritsenko 1995]. The action of $T(q)$ on the Fourier–Jacobi expansion of an $f \in S_k(K(N))$ is given by

$$\mathrm{FJ}(f) = \sum_{j=1}^{\infty} \phi_j \xi^{Nj}; \quad \mathrm{FJ}(f|T(q)) = \sum_{j=1}^{\infty} \left(\phi_{qj}|W_q + q^{k-2}\phi_{j/q}|V_q\right) \xi^{Nj}, \tag{27}$$

as can be directly verified by comparing equations (25) and (26) with (17) using (20).

3.4. ***Jacobi restriction and upper bounds.*** In this section we define the Jacobi restriction spaces $\mathcal{J}_u^\epsilon(R)$ for R being $\mathbb{F}_p$ or a subring of $\mathbb{C}$. Jacobi restriction is described in [Ibukiyama et al. 2013; Breeding et al. 2016] but we cover it here in further detail because the extension of $T(q)$ to $\mathcal{J}_u^\epsilon(\mathbb{F}_p)$ in Section 3.7 is subtle.

By collectively ordering the index sets of the Fourier expansions of $J_{k,Nj}^{\mathrm{cusp}}$ for all $j \in \mathbb{N}$ in some way, we view $\prod_{j=1}^{\infty} J_{k,Nj}^{\mathrm{cusp}}(R) \subseteq R^\infty$.

Definition 3.4.1. Let $N, u, D_0 \in \mathbb{N}$, $k \in \mathbb{Z}$, and $\epsilon \in \{-1, 1\}$. Let R be $\mathbb{F}_p$ or a subring of $\mathbb{C}$. The R-module

$$\mathcal{J}_u^\epsilon(R) \subseteq \bigoplus_{j=1}^{u} J_{k,Nj}^{\mathrm{cusp}}(R) \subseteq R^\infty$$

consists of the $\mathfrak{f} = \sum_{j=1}^{u} \mathfrak{f}_j\, \xi^{Nj} \in \bigoplus_{j=1}^{u} J_{k,Nj}^{\mathrm{cusp}}(R)$ that satisfy the following conditions:

$$\begin{gathered}\text{for all } t_1 = \left[\begin{smallmatrix} n_1 & r_1/2 \\ r_1/2 & Nm_1 \end{smallmatrix}\right],\ t_2 = \left[\begin{smallmatrix} n_2 & r_2/2 \\ r_2/2 & Nm_2 \end{smallmatrix}\right] \in \mathcal{X}_2(N) \text{ and } U \in \hat{\Gamma}^0(N),\\ t_1[U] = t_2 \text{ and } \det(2t_1), \det(2t_2) \le D_0 \text{ and } m_1, m_2 \le u \implies c(n_1, r_1; \mathfrak{f}_{m_1}) = \det(U)^k c(n_2, r_2; \mathfrak{f}_{m_2}),\end{gathered} \tag{28}$$

and

$$\text{for all } t = \left[\begin{smallmatrix} n & r/2 \\ r/2 & Nm \end{smallmatrix}\right] \in \mathcal{X}_2(N), \quad \det(2t) \le D_0 \text{ and } n, m \le u \implies c(n, r; \mathfrak{f}_m) = (-1)^k \epsilon\, c(m, r; \mathfrak{f}_n). \tag{29}$$

This important construction calls for a number of comments. The defining equations in Definition 3.4.1 are truly elementary, one coordinate in R^∞ equals ± 1 times another, so that $\mathcal{J}_u^\epsilon(R)$ is defined over the various commutative rings R. The R-module $\mathcal{J}_u^\epsilon(R)$ also depends on N, k, and D_0 so that $\mathcal{J}_u^\epsilon(R, N, k, D_0)$ would be more proper, but we supress N, k, and D_0 to lighten the notation somewhat. When no ring is indicated the field of complex numbers is meant, so $\mathcal{J}_u^\epsilon = \mathcal{J}_u^\epsilon(\mathbb{C})$. We have written a program, which we call Jacobi restriction, for the cases $R = \mathbb{Z}$ and $R = \mathbb{F}_p$. This program accepts input $(N, k, \epsilon, D_0, u, R)$ and returns initial expansions, out to (n, r) satisfying $4nNj - r^2 \le D_0$, of an R-basis of $\mathcal{J}_u^\epsilon(R)$. We always choose D_0 large enough so that elements of $J_{k,Nj}^{\mathrm{cusp}}(R)$ for $j \le u$ are determined by their initial expansions out to $4nNj - r^2 \le D_0$; thus, the output characterizes a basis of $\mathcal{J}_u^\epsilon(R)$, and $\mathcal{J}_u^\epsilon(R)$ is an R-module of finite rank very amenable to computation. In particular, $\mathrm{rank}_R\, \mathcal{J}_u^\epsilon(R)$ is always known. Finally, because the spaces $J_{k,m}^{\mathrm{cusp}}$ have integral bases, the output for $R = \mathbb{Z}$ also works for any subring $R \subseteq \mathbb{C}$.

The next lemma shows that $\mathcal{J}_u^\epsilon$ is an upper approximation of the space $S_k(K(N))^\epsilon[u]$.

Lemma 3.4.2. *Let $N, u \in \mathbb{N}$, $k \in \mathbb{Z}$, and $\epsilon \in \{-1, 1\}$. We have*

$$\operatorname{proj}_u^\infty \circ \mathrm{FJ} : S_k(K(N))^\epsilon \to S_k(K(N))^\epsilon[u] \subseteq \mathcal{J}_u^\epsilon.$$

Proof. By equations (23) and (24), the Fourier–Jacobi expansion of an $f \in S_k(K(N))^\epsilon$ satisfies the conditions in Definition 3.4.1 for all choices of indices. The conditions defining $\mathcal{J}_u^\epsilon$ are thus a subset of the conditions satisfied by $(\operatorname{proj}_u^\infty \circ \mathrm{FJ})(f)$. □

Corollary 3.4.3. *Let $u \in \mathbb{N}$ be such that $\operatorname{proj}_u^\infty \circ \mathrm{FJ} : S_k(K(N))^\epsilon \to S_k(K(N))^\epsilon[u]$ injects. Then $\dim S_k(K(N))^\epsilon \le \dim \mathcal{J}_u^\epsilon$.*

3.5. *Jacobi restriction modulo p.* Jacobi restriction can also be run modulo a prime p. As in the appendix of [Berger and Klosin 2017], for a subset $H \subseteq \mathbb{C}^\infty$, let $H_p = \mathrm{R_p}\,(H \cap \mathbb{Z}^\infty) \subseteq \mathbb{F}_p^\infty$ denote the reduction of $H \cap \mathbb{Z}^\infty$ mod p. If $H_1, H_2 \subseteq \mathbb{C}^\infty$ are subspaces with integral bases and $L : H_1 \to H_2$ is a linear map whose matrix in these bases is integral, then L also has a reduction, $L_p : H_{1p} \to H_{2p}$, with the defining property that $(L(h))_p = L_p(h_p)$ for $h \in H_1$. To give some examples, for paramodular forms we have $(\mathrm{FE}(S_k(K(N))))_p = S_k(K(N))(\mathbb{F}_p)$ and for Jacobi forms $(\mathrm{FE}(J_{k,m}^{\mathrm{cusp}}))_p = J_{k,m}^{\mathrm{cusp}}(\mathbb{F}_p)$. The good Hecke operator $T(q) : S_k(K(N))^\epsilon(\mathbb{Z}) \to S_k(K(N))^\epsilon(\mathbb{Z})$ has, for $k \ge 2$, an integral matrix by (17), and so induces a map $T(q)_p : S_k(K(N))^\epsilon(\mathbb{F}_p) \to S_k(K(N))^\epsilon(\mathbb{F}_p)$ given by: $\mathfrak{f}|T(q)_p = \mathfrak{g}$ means there exists an $f \in S_k(K(N))^\epsilon(\mathbb{Z})$ such that $\mathrm{R_p}\,(\mathrm{FE}(f)) = \mathfrak{f}$ and $\mathrm{R_p}\,(\mathrm{FE}(f|T(q))) = \mathfrak{g}$.

Because spaces of modular forms have integral bases, important information survives the reduction mod p. For example, $\dim_{\mathbb{C}} S_k(K(N))^\epsilon[u] = \dim_{\mathbb{F}_p} S_k(K(N))^\epsilon[u]_p \le \dim \mathcal{J}_{u,p}^\epsilon$. Hence if $u \ge u_0$, for some basic u_0 making $\operatorname{proj}_{u_0}^\infty \circ \mathrm{FJ}$ injective, we have $\dim_{\mathbb{C}} S_k(K(N))^\epsilon \le \dim \mathcal{J}_{u,p}^\epsilon$ as well. We easily have $\mathcal{J}_{u,p}^\epsilon \subseteq \mathcal{J}_u^\epsilon(\mathbb{F}_p)$ and examples show that the containment can be proper. Noting Lemma 3.4.2, the *hope* when we run Jacobi restriction is that all the following spaces have the same dimension:

$$S_k(K(N))^\epsilon \xrightarrow{\operatorname{proj}_u^\infty \circ \mathrm{FJ}} S_k(K(N))^\epsilon[u] \xrightarrow{\bmod p} S_k(K(N))^\epsilon[u]_p \subseteq \mathcal{J}_{u,p}^\epsilon \subseteq \mathcal{J}_u^\epsilon(\mathbb{F}_p). \tag{30}$$

When these spaces do have the same dimension we can, in retrospect, regard the computations as having been perfomed in any one of them; however it is the space $\mathcal{J}_u^\epsilon(\mathbb{F}_p)$ that is most amenable to computation, being a finite-dimensional $\mathbb{F}_p$-vector space with a known basis. Especially, we can row reduce and compute the smallest u_1^ϵ for which the projection

$$\operatorname{proj}_{u_1^\epsilon}^u : \mathcal{J}_u^\epsilon(\mathbb{F}_p) \to \bigoplus_{j=1}^{u_1^\epsilon} J_{k,Nj}^{\mathrm{cusp}}(\mathbb{F}_p)$$

is injective. For $u = u_0$, Table 3 also gives particular values of u_1^ϵ with this property for $1 \le k \le 14$, $N \in \{4, 8, 16\}$, $p = 12347$, and various D_0. The choice of D_0 was 400 for $K(4)$, 800 for $K(8)$ when $k \le 10$ and 1000 for larger k, and 1600 for $K(16)$ when $k \le 10$ and 2000 for larger k. The caption of Table 3, however, instead reports that the projection from $S_k(K(N))^\epsilon$ to $S_k(K(N))^\epsilon[u_1^\epsilon]$ is injective. The injectivity in these cases follows from the proof in Section 3.10 that $\dim S_k(K(N))^\epsilon = \dim \mathcal{J}_{u_1^\epsilon}^\epsilon(\mathbb{F}_p)$, and so p and D_0 are not reported in Table 3.

3.6. *Extending $T(q)$ to $\mathcal{J}_u^\epsilon(\mathbb{C})$.* Our goal in this section is to lift the map $T(q): S_k(K(N))^\epsilon \to S_k(K(N))^\epsilon$ to another map $\hat{T}(q): \mathcal{J}_u^\epsilon \to \mathcal{J}_u^\epsilon$ such that the following diagram commutes:

$$\begin{tikzcd}
\mathcal{J}_u^\epsilon \arrow[r, "\hat{T}(q)"] & \mathcal{J}_u^\epsilon \\
S_k(K(N))^\epsilon \arrow[u, "\mathrm{proj}_u^\infty \circ \mathrm{FJ}"] \arrow[r, "T(q)"] & S_k(K(N))^\epsilon \arrow[u, "\mathrm{proj}_u^\infty \circ \mathrm{FJ}"']
\end{tikzcd} \tag{31}$$

Admittedly, this diagram will only be useful for u large enough to make the vertical map injective. We proceed in two steps and need to make certain assumptions about the space $\mathcal{J}_u^\epsilon$. Because we can compute with $\mathcal{J}_u^\epsilon$ it is reasonable to impose needed conditions on $\mathcal{J}_u^\epsilon$ as long as they can be checked in practice. First, define a map

$$\tilde{T}(q): \bigoplus_{j=1}^{u} J_{k,Nj}^{\mathrm{cusp}} \to \bigoplus_{j=1}^{\lfloor u/q \rfloor} J_{k,Nj}^{\mathrm{cusp}}, \qquad \sum_{j=1}^{u} \phi_j\, \xi^{Nj} \mapsto \sum_{j=1}^{\lfloor u/q \rfloor} \left(q^{k-2}\phi_{j/q}|V_q + \phi_{qj}|W_q\right) \xi^{Nj}. \tag{32}$$

This definition reflects the computational fact that the operator $T(q)$ returns shorter Fourier–Jacobi expansions than it receives. Since the above action agrees with equation (27) we have

$$\mathrm{proj}^u_{\lfloor u/q \rfloor}\, \mathrm{proj}^\infty_u\, \mathrm{FJ}(f|T(q)) = \left(\mathrm{proj}^\infty_u\, \mathrm{FJ}(f)\right) |\tilde{T}(q).$$

We introduce the notion of one map being relatively stable with respect to another. Let $\pi: A \to \pi A$ and $T: A \to \pi A$ be maps and $B \subseteq A$. We say T is *relatively stable on B with respect to π* when $T(B) \subseteq \pi(B)$. This is equivalent to saying that $T: A \to \pi A$ extends to a relative map $T: (A, B) \to \pi(A, B)$. We will require that $\tilde{T}(q)$ be relatively stable on $\mathcal{J}_u^\epsilon$ with respect to $\mathrm{proj}^u_{\lfloor u/q \rfloor}$. When $\mathcal{J}_u^\epsilon$ has successfully been computed, we will need to check whether or not $\tilde{T}(q): \mathcal{J}_u^\epsilon \to \mathrm{proj}^u_{\lfloor u/q \rfloor}\, \mathcal{J}_u^\epsilon \subseteq \bigoplus_{j=1}^{\lfloor u/q \rfloor} J_{k,Nj}^{\mathrm{cusp}}$. We will also require that $\lfloor u/q \rfloor \ge u_1^\epsilon$, so that $\mathrm{proj}^u_{\lfloor u/q \rfloor}$ injects on $\mathcal{J}_u^\epsilon$.

Proposition 3.6.1. *Let $N, u \in \mathbb{N}$, $k \in \mathbb{Z}$, and $\epsilon \in \{-1, 1\}$. Let q be a prime with $q \nmid N$. Assume that*:

i) *$\tilde{T}(q)$ is relatively stable on $\mathcal{J}_u^\epsilon$ with respect to* $\mathrm{proj}^u_{\lfloor u/q \rfloor}$.

ii) *The restriction of* $\mathrm{proj}^u_{\lfloor u/q \rfloor}$ *to $\mathcal{J}_u^\epsilon$ is injective.*

Then $\hat{T}(q): \mathcal{J}_u^\epsilon \to \mathcal{J}_u^\epsilon$ is well-defined by: $\mathfrak{f}|\hat{T}(q) = \mathfrak{g}$ means $\mathfrak{f}|\tilde{T}(q) = \mathrm{proj}^u_{\lfloor u/q \rfloor}\, \mathfrak{g}$. Under these hypotheses, diagram (31) commutes.

Proof. Assume that $\mathfrak{f} \in \mathcal{J}_u^\epsilon$. Because $\tilde{T}(q)$ is relatively stable there exists a $\mathfrak{g} \in \mathcal{J}_u^\epsilon$ such that $\mathfrak{f}|\tilde{T}(q) = \mathrm{proj}^u_{\lfloor u/q \rfloor}\, \mathfrak{g}$. Because $\mathrm{proj}^u_{\lfloor u/q \rfloor}$ is injective, this $\mathfrak{g}$ is unique, and thus $\hat{T}(q)$ is well-defined. The linearity of $\hat{T}(q)$ follows from the equation $\mathfrak{f}|\tilde{T}(q) = \mathrm{proj}^u_{\lfloor u/q \rfloor}\, \mathfrak{g}$ and the uniqueness of $\mathfrak{g}$.

In order to show the commutativity of the diagram we must check

$$\left(\mathrm{proj}^\infty_u(\mathrm{FJ}(f))\right) |\hat{T}(q) = (\mathrm{proj}^\infty_u \circ \mathrm{FJ})(f|T(q)),$$

or, by definition of $\hat{T}(q)$, we must show that

$$\left(\mathrm{proj}^\infty_u(\mathrm{FJ}(f))\right) |\tilde{T}(q) = \mathrm{proj}^u_{\lfloor u/q \rfloor} \left((\mathrm{proj}^\infty_u \circ \mathrm{FJ})(f|T(q))\right).$$

Thus we must check that $\left(\sum_{j=1}^{u} \phi_j\, \xi^{Nj}\right)|\tilde{T}(q) = \mathrm{proj}^{\infty}_{\lfloor u/q\rfloor}(f|T(q))$. By equation (27) the right-hand side is $\sum_{j=1}^{\lfloor u/q\rfloor} \left(q^{k-2}\phi_{j/q}|V_q + \phi_{qj}|W_q\right)\xi^{Nj}$, which is the definition of the left-hand side. □

3.7. *Extending $T(q)_p$ to $\mathcal{J}_u^{\epsilon}(\mathbb{F}_p)$.* Our goal in this section is to lift the map $T(q)_p : S_k(K(N))^{\epsilon}(\mathbb{F}_p) \to S_k(K(N))^{\epsilon}(\mathbb{F}_p)$ to a map $\mathcal{T}(q) : \mathcal{J}_u^{\epsilon}(\mathbb{F}_p) \to \mathcal{J}_u^{\epsilon}(\mathbb{F}_p)$ such that the following diagram commutes:

$$\begin{tikzcd}
\mathcal{J}_u^{\epsilon}(\mathbb{F}_p) \arrow[r, "\mathcal{T}(q)"] & \mathcal{J}_u^{\epsilon}(\mathbb{F}_p) \\
S_k(K(N))^{\epsilon}(\mathbb{F}_p) \arrow[u, "\mathrm{proj}^{\infty}_{u,p}\circ \mathrm{FJ}_p"] \arrow[r, "T(q)_p"] & S_k(K(N))^{\epsilon}(\mathbb{F}_p)\,. \arrow[u, "\mathrm{proj}^{\infty}_{u,p}\circ \mathrm{FJ}_p"']
\end{tikzcd} \tag{33}$$

Recall the definition (32) of the map $\tilde{T}(q)$. By equations (25) and (26), the action of V_q and W_q is integral for $k \geq 2$; so we may consider the reduction of the map $\tilde{T}(q)$ mod p:

$$\tilde{T}(q)_p : \bigoplus_{j=1}^{u} J^{\mathrm{cusp}}_{k,Nj}(\mathbb{F}_p) \to \bigoplus_{j=1}^{\lfloor u/q\rfloor} J^{\mathrm{cusp}}_{k,Nj}(\mathbb{F}_p),$$

and restrict $\tilde{T}(q)_p$ to $\mathcal{J}_u^{\epsilon}(\mathbb{F}_p) \subseteq \bigoplus_{j=1}^{u} J^{\mathrm{cusp}}_{k,Nj}(\mathbb{F}_p)$ to obtain $\tilde{T}(q)_p : \mathcal{J}_u^{\epsilon}(\mathbb{F}_p) \to \bigoplus_{j=1}^{\lfloor u/q\rfloor} J^{\mathrm{cusp}}_{k,Nj}(\mathbb{F}_p)$. As in the previous section, it is reasonable to impose needed conditions on $\mathcal{J}_u^{\epsilon}(\mathbb{F}_p)$ that are easy to check. We will require that $\tilde{T}(q)_p$ be relatively stable on $\mathcal{J}_u^{\epsilon}(\mathbb{F}_p)$ with respect to $\mathrm{proj}^{u}_{\lfloor u/q\rfloor,p}$. This condition is achieved whenever $S_k(K(N))^{\epsilon}[u]_p$ actually equals $\mathcal{J}_u^{\epsilon}(\mathbb{F}_p)$, which is what the whole set-up aims to prove, so there is no harm in requiring relative stability. If relative stability fails, we should increase u and try again. We will also require that $\mathrm{proj}^{u}_{\lfloor u/q\rfloor,p}$ be injective on $\mathcal{J}_u^{\epsilon}(\mathbb{F}_p)$. This second condition is achieved when $\lfloor u/q\rfloor \geq u_1^{\epsilon}$, which may be costly.

Proposition 3.7.1. *Let $N, u, k \in \mathbb{N}$ with $k \geq 2$, and $\epsilon \in \{-1, 1\}$. Let p and q be primes with $q \nmid N$. Assume that*:

i) *$\tilde{T}(q)_p$ is relatively stable on $\mathcal{J}_u^{\epsilon}(\mathbb{F}_p)$ with respect to $\mathrm{proj}^{u}_{\lfloor u/q\rfloor,p}$.*

ii) *The restriction of $\mathrm{proj}^{u}_{\lfloor u/q\rfloor,p}$ to $\mathcal{J}_u^{\epsilon}(\mathbb{F}_p)$ is injective.*

Then $\mathcal{T}(q) : \mathcal{J}_u^{\epsilon}(\mathbb{F}_p) \to \mathcal{J}_u^{\epsilon}(\mathbb{F}_p)$ is well-defined by: $\mathfrak{f}|\mathcal{T}(q) = \mathfrak{g}$ means $\mathfrak{f}|\tilde{T}(q)_p = \mathrm{proj}^{u}_{\lfloor u/q\rfloor,p}\,\mathfrak{g}$. Under these hypotheses, diagram (33) commutes.

Proof. We show that $\mathcal{T}(q)$ is well-defined and $\mathbb{F}_p$-linear. Take $\mathfrak{f} \in \mathcal{J}_u^{\epsilon}(\mathbb{F}_p)$. Since $\tilde{T}(q)_p$ is relatively stable on $\mathcal{J}_u^{\epsilon}(\mathbb{F}_p)$ with respect to $\mathrm{proj}^{u}_{\lfloor u/q\rfloor,p}$, there exists a $\mathfrak{g} \in \mathcal{J}_u^{\epsilon}(\mathbb{F}_p)$ such that $\mathfrak{f}|\tilde{T}(q)_p = \mathrm{proj}^{u}_{\lfloor u/q\rfloor,p}\,\mathfrak{g}$. If there were another such $\mathfrak{g}'$, then $\mathfrak{g}' = \mathfrak{g}$ because $\mathrm{proj}^{u}_{\lfloor u/q\rfloor,p}$ is injective on $\mathcal{J}_u^{\epsilon}(\mathbb{F}_p)$. This shows that $\mathcal{T}(q)$ is well-defined. Linearity follows from $\mathfrak{f}|\tilde{T}(q)_p = \mathrm{proj}^{u}_{\lfloor u/q\rfloor,p}\,\mathfrak{g}$ and the uniqueness of $\mathfrak{g}$.

In order to show the commutativity of the diagram, take $\mathfrak{f} \in S_k(K(N))^{\epsilon}(\mathbb{F}_p)$. We must show

$$\left(\mathrm{proj}^{\infty}_{u,p}(\mathrm{FJ}_p(\mathfrak{f}))\right)|\mathcal{T}(q) = (\mathrm{proj}^{\infty}_{u,p}\circ \mathrm{FJ}_p)(\mathfrak{f}|T(q)_p),$$

which, by definition of $\mathcal{T}(q)$, means

$$\left(\mathrm{proj}^{\infty}_{u,p}(\mathrm{FJ}_p(\mathfrak{f}))\right)|\tilde{T}(q)_p = \mathrm{proj}^{u}_{\lfloor u/q\rfloor,p}\left((\mathrm{proj}^{\infty}_{u,p}\circ \mathrm{FJ}_p)(\mathfrak{f}|T(q)_p)\right),$$

or equivalently,

$$\big(\mathrm{proj}^{\infty}_{u,p}(\mathrm{FJ}_p(\mathfrak{f}))\big)|\tilde{T}(q)_p = \mathrm{proj}^{u}_{\lfloor u/q\rfloor,p}\,\mathrm{FJ}_p(\mathfrak{f}|T(q)_p). \tag{34}$$

There is an $f \in S_k(K(N))^{\epsilon}(\mathbb{Z})$ such that $\mathfrak{f} = \mathrm{FJ}(f)_p$, so that (34) would follow by reduction from

$$\big(\mathrm{proj}^{\infty}_{u}(\mathrm{FJ}(f))\big)\,|\tilde{T}(q) = \mathrm{proj}^{u}_{\lfloor u/q\rfloor}\,\mathrm{FJ}(f|T(q)). \tag{35}$$

Writing $\mathrm{FJ}(f) = \sum_{j=1}^{\infty}\phi_j\,\xi^{Nj}$, we verify (35) from the definition of $\tilde{T}(q)$ and equation (27),

$$\Big(\sum_{j=1}^{u}\phi_j\,\xi^{Nj}\Big)|\tilde{T}(q) = \sum_{j=1}^{\lfloor u/q\rfloor}\big(q^{k-2}\phi_{j/q}|V_q + \phi_{qj}|W_q\big)\,\xi^{Nj} = \sum_{j=1}^{\lfloor u/q\rfloor}\phi_j(f|T(q))\,\xi^{Nj}. \qquad \square$$

3.8. ***Bootstrapping and lower bounds.*** We now explain the technique of *bootstrapping*, a combination of Jacobi restriction and Hecke spreading, which computes lower bounds for $\dim S_k(K(N))^{\epsilon} = \dim S_k(K(N))^{\epsilon}(\mathbb{F}_p)$. As motivation, we first discuss Borcherds products. The theory of Borcherds products and the theory of Hecke operators bear little relation. A Borcherds product, for example, seems to only be a Hecke eigenform when forced to be by dimensional reasons. In general, if a Borcherds product is written as a linear combination of Hecke eigenforms it seems that the Borcherds product is often supported on every eigenspace with the same Atkin–Lehner signs as the Borcherds product. Thus repeated applications of $T(q)$ on a Borcherds product are likely to span the entire Atkin–Lehner space of paramodular forms that the Borcherds product belongs to. Over $\mathbb{Q}$, many iterations of $T(q)$ on a Borcherds product are much too expensive, but over $\mathbb{F}_p$ many iterations of $\mathcal{T}(q)$ on $\mathcal{J}^{\epsilon}_u(\mathbb{F}_p)$ are feasible.

Let $S \subseteq S_k(K(N))^{\epsilon}(\mathbb{F}_p)$. Define

$$B_p(S;\mathcal{T}(q)) = \mathrm{Span}_{\mathbb{F}_p}\big\{(\mathrm{proj}^{\infty}_{u,p}\circ\mathrm{FJ}_p(\mathfrak{f}))|\mathcal{T}(q)^i \in \mathcal{J}^{\epsilon}_u(\mathbb{F}_p) : i \in \mathbb{Z}_{\geq 0}, \mathfrak{f} \in S\big\}.$$

Lemma 3.8.1. *Let u be large enough so that $\mathrm{proj}^{\infty}_{u,p}\circ\mathrm{FJ}_p$ injects on $S_k(K(N))^{\epsilon}(\mathbb{F}_p)$. Assume the hypotheses of Proposition 3.7.1. Then*

$$\dim B_p(S;\mathcal{T}(q)) \leq \dim S_k(K(N))^{\epsilon}(\mathbb{F}_p).$$

Proof. By the commutative diagram (33), the subspace $B_p(S;\mathcal{T}(q)) \subseteq \mathcal{J}^{\epsilon}_u(\mathbb{F}_p)$ is the injective image under $\mathrm{proj}^{\infty}_{u,p}\circ\mathrm{FJ}_p$ of the span of $\mathfrak{f}|T(q)^i_p \in S_k(K(N))^{\epsilon}(\mathbb{F}_p)$ for $i \in \mathbb{Z}_{\geq 0}$, and $\mathfrak{f} \in S$. $\square$

3.9. ***Specific upper bounds: Jacobi restriction.*** We use the technique of Jacobi restriction to compute upper bounds for $\dim S_k(K(N))^{\epsilon}$. Jacobi restriction over $\mathbb{Q}$ requires a lot of memory. It is better, when sufficient, to run Jacobi restriction modulo p. Table 3 gives u_0 large enough to make projection onto the first u_0 Jacobi coefficients injective. Using the containments in (30), Table 4 reports the resulting upper bound $\dim S_k(K(N))^{\epsilon} = \dim S_k(K(N))^{\epsilon}[u_0] \leq \dim \mathcal{J}^{\epsilon}_{u_0}(\mathbb{F}_p)$ given as output by the Jacobi restriction program, using the same determinant bounds D_0 and prime p as in Section 3.5. In Table 4 we have further refined these upper bounds to apply to the spaces of nonlifts, which is a direct adjustment because the dimensions of the lift spaces are known by [Eichler and Zagier 1985]. Because $\dim S_k(K(4))$ is known and the upper bounds for the three subspaces of $S_k(K(4))$ add up to the known total dimension, the dimensions of the subspaces of $S_k(K(4))$ listed in Table 4 are the actual dimensions without further argument. We will prove that the upper bounds of the dimensions of the nonlift subspaces of $S_k(K(8))$ and $S_k(K(16))$ as listed in Table 4 are in fact the true dimensions in all cases. This illustrates the power

	$K(1)$			$K(2)$			$K(4)$			$K(8)$			$K(16)$		
k	lifts	nonlifts		lifts	nonlifts		lifts	nonlifts		lifts	nonlifts		lifts	nonlifts	
		$+$	$-$		$+$	$-$		$+$	$-$		$+$	$-$		$+$	$-$
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
5	0	0	0	0	0	0	0	0	0	1	0	0	2	0	0
6	0	0	0	0	0	0	0	0	0	1	0	0	3	1	0
7	0	0	0	0	0	0	1	0	0	2	0	0	5	0	2
8	0	0	0	1	0	0	1	0	0	3	0	0	6	5	0
9	0	0	0	0	0	0	1	0	0	3	0	1	7	1	8
10	1	0	0	1	0	0	2	0	0	4	2	0	9	13	2
11	0	0	0	1	0	0	2	0	1	5	1	3	10	4	19
12	1	0	0	2	0	0	3	1	0	6	5	1	12	27	6
13	0	0	0	0	0	0	2	0	1	5	2	6	12	10	34
14	1	0	0	2	0	0	3	2	0	7	9	3	14	46	14

Table 4. Dimensions of cusp forms of weight k. The signs $+$ and $-$ refer to the paramodular Atkin–Lehner sign, which is the same as the Fricke sign in these cases.

of Jacobi restriction. The proof involves constructing enough paramodular forms to show these numbers are also lower bounds.

3.10. ***Specific lower bounds: Borcherds products and bootstrapping.*** In the previous section we computed the upper bounds for dim $S_k(K(N))$ given in Table 4. This section will compute matching lower bounds, mainly by constructing Gritsenko lifts and Borcherds products, but also via Hecke operators, and oldform theory. The theory of Borcherds products [Borcherds 1998; Gritsenko and Nikulin 1998] creates meromorphic paramodular forms, transforming by a character χ of $K(N)$, in $M_k^{\text{mero}}(K(N))^{\epsilon}(\chi)$ from weakly holomorphic Jacobi forms $\psi \in J_{0,N}^{\text{wh}}$ of weight zero and index N whose Fourier coefficients are integral on singular indices. We will only use Borcherds products that turn out to be holomorphic and cuspidal with trivial character. There is an algorithm [Poor et al. 2018] to find all Borcherds products in a given space $S_k(K(N))$, so we simply post the constructions of the Borcherds products that we use here on the website [Yuen 2018]. Given an appropriate $\psi \in J_{0,N}^{\text{wh}}$, we write $\text{Borch}(\psi) \in M_k^{\text{mero}}(K(N))^{\epsilon}(\chi)$ for the associated Borcherds product. If we write the Fourier expansion of ψ as $\psi(\tau, z) = \sum_{n,r\in\mathbb{Z}} c(n,r) q^n \zeta^r$, then $\text{Borch}(\psi)$ is defined by analytic continuation of the following infinite product for $\Omega = \left[\begin{smallmatrix}\tau & z\\ z & \omega\end{smallmatrix}\right] \in \mathcal{H}_2$:

$$\text{Borch}(\psi)(\Omega) = q^A \zeta^B \xi^C \prod_{(m,n,r)\geq 0} (1 - q^n \zeta^r \xi^{mN})^{c(nm,r)}.$$

The product is taken over $m, n, r \in \mathbb{Z}$ such that $m \geq 0$, and if $m = 0$ then $n \geq 0$, and if $m = n = 0$ then $r < 0$. Set $\mathbb{N} = \{1, 2, 3, \dots\}$. The exponents A, B, C are given by $24A = \sum_{r\in\mathbb{Z}} c(0,r)$, $2B = \sum_{r\in\mathbb{N}} r c(0,r)$, and $2C = \sum_{r\in\mathbb{N}} r^2 c(0,r)$. Borcherds products always come with a Fricke sign. The sign ϵ is given by $\epsilon = (-1)^{d_o}$ where $d_o = \sum_{n\in\mathbb{N}} \sigma_0(n)c(-n,0)$, and $\sigma_0(n)$ is the number of positive divisors of n.

Here are our methods for obtaining lower bounds on $\dim S_k(K(N))^\epsilon$. Fix k, N, and $\epsilon = \pm 1$. We search for Borcherds products in $S_k(K(N))^\epsilon$. If we find enough to span a space whose dimension equals that of the upper bound, then we are done. If not, we employ the method of bootstrapping from Section 3.8. We check the hypotheses of Proposition 3.7.1: that $\tilde{T}(3)_p$ is relatively stable on $\mathcal{J}^\epsilon_{u_0}(\mathbb{F}_p)$ with respect to $\mathrm{proj}^{u_0}_{\lfloor u_0/3\rfloor}$, and that $u_0 \geq 3u^\epsilon_1$ so that $\mathrm{proj}^{u_0}_{\lfloor u_0/3\rfloor}$ is injective on $\mathcal{J}^\epsilon_{u_0}(\mathbb{F}_p)$. There are three places in Table 3 where $u_0 < 3u^\epsilon_1$, but these occur for $K(4)$ and weight $k \in \{7, 11, 12\}$ where the dimension is already known. Still using the u_0 from Table 3, we compute a matrix representation for $\mathcal{T}(3)$ on a fixed basis for $\mathcal{J}^\epsilon_{u_0}(\mathbb{F}_p)$. We find a set $\tilde{S} \subseteq S_k(K(N))^\epsilon$ of Borcherds products and take $f \in \tilde{S}$; see [Yuen 2018] for the Borcherds products found. It is feasible to expand a Borcherds product f out far enough to determine $(\mathrm{proj}^\infty_{u_0} \mathrm{FJ}(f))_p$ in this basis. Define $S = (\mathrm{FJ}(\tilde{S}))_p \subseteq S_k(K(N))^\epsilon(\mathbb{F}_p)$. Once we get the coordinates of $(\mathrm{proj}^\infty_{u_0} \mathrm{FJ}(f))_p$ in this basis, it is linear algebra to compute the bootstrapped subspace on S. Then $u_0 \geq 3u^\epsilon_1$ and Lemma 3.8.1 imply that $\dim B_p(S; \mathcal{T}(3)) \leq \dim S_k(K(N))^\epsilon(\mathbb{F}_p)$. It turns out that the dimension of each bootstrapped subspace $B_p(S; \mathcal{T}(3))$ gives the same lower bound as the upper bound $\dim \mathcal{J}^\epsilon_{u_0}(\mathbb{F}_p)$ in every case in Table 4 except in the single case $S_{14}(K(8))^-$. Thus we know $\dim_{\mathbb{C}} S_k(K(N))^\epsilon = \dim_{\mathbb{C}} S_k(K(N))^\epsilon[u_0] = \dim_{\mathbb{F}_p} S_k(K(N))^\epsilon[u_0]_p = \dim_{\mathbb{F}_p} \mathcal{J}^\epsilon_{u_0}(\mathbb{F}_p)$ in all cases in Table 4 except $S_{14}(K(8))^-$. There are no Borcherds products in $S_{14}(K(8))^-$. We now explain the additional argument needed for this exceptional case.

We know that $\dim S_{14}(K(8))^- \leq 3$. We found all the eigenforms in each of $S_{14}(K(N))^\pm$ for $N \in \{1, 2, 4, 8, 16\}$ except $S_{14}(K(8))^-$. We show there is an eigenform in $S_{14}(K(16))^-$ of $T(3)$-eigenvalue $3^{11}\lambda_3 = -1580472$ which is not a $T_{1,0}$-eigenform. The eigenspace of $S_{14}(K(16))^-$ with this $T(3)$-eigenvalue is one-dimensional. Lemma 3.10.1 implies that there exists a newform $f_{\mathrm{new}} \in S_{14}(K(2^j))$ for some $j \in \{0, 1, 2, 3\}$ with the same $T(3)$-eigenvalue. Looking at $T(3)$-eigenvalues for the lifts, we see that f_{new} must be a nonlift. There are no nonlifts in $S_{14}(K(N))$ for $N \in \{1, 2\}$ and there are two nonlift eigenforms in $S_{14}(K(4))$. But the $T(3)$-eigenvalue -1580472 does not show as an eigenvalue in $S_{14}(K(8))^+$ or in $S_{14}(K(4))$. We conclude that f_{new} must be in $S_{14}(K(8))^-$. Together with the two oldforms in $S_{14}(K(8))^-$ coming from the two newforms in $S_{14}(K(4))$, we conclude that $\dim S_{14}(K(8))^- \geq 2+1 = 3$.

Lemma 3.10.1. *Let N be a positive integer and p be a prime dividing N. Let $W \subset S_k(K(N))$ be a nonzero eigenspace for a Hecke operator T at some good place $q \nmid N$. Assume that the operators $T_{0,1}(p)$, $T_{1,0}(p)$ and the Atkin–Lehner α_p are not simultaneously diagonalizable on W. Then there exists a new-eigenform $f_{\mathrm{new}} \in S_k(K(M))$ for some $M|N$ with $v_p(M) < v_p(N)$ and with the same T-eigenvalue as the elements of W.*

Proof. Since Hecke operators at good places commute, we can find a basis $f_1, \ldots, f_n$ of W consisting of eigenforms for almost all good Hecke operators, including the place q. By Theorem 2.6 i) of [Schmidt 2018], the adelization Φ_i of f_i generates an irreducible, cuspidal, automorphic representation $\pi_i \cong \bigotimes_s \pi_{i,s}$ of $\mathrm{PGSp}(4, \mathbb{A}_{\mathbb{Q}})$, for each i. The automorphic form Φ_i corresponds to a sum of pure tensors $\sum_j \big(\bigotimes_s w_{i,s,j}\big)$, where $w_{i,s,j}$ is in the space of $\pi_{i,s}$. After averaging, we may assume that $w_{i,s,j}$ is a paramodular vector of level $v_s(N)$, for each prime number s. In particular, each $w_{i,q,j}$ is a spherical vector in $\pi_{i,q}$, and hence an eigenvector for the local operator T_q corresponding to T, with the same eigenvalue as T on W.

We claim that there exists an $i \in \{1, \dots, n\}$ such that the conductor exponent $a(\pi_{i,p})$ is less than $v_p(N)$. Clearly, we must have $a(\pi_{i,p}) \le v_p(N)$ for each i, since $a(\pi_{i,p})$ is the smallest possible level of any paramodular vector in $\pi_{i,p}$ by Corollary 7.5.5 of [Roberts and Schmidt 2007]. Assume that we would have $a(\pi_{i,p}) = v_p(N)$ for all i. Then each $w_{i,p,j}$ would be a local newform in $\pi_{i,p}$, which is unique up to scalars by Theorem 7.5.4 of the same reference. In particular, $T_{0,1}(p)$, $T_{1,0}(p)$ and α_p would be simultaneously diagonalizable on W, contradicting our hypothesis. This proves our claim that there exists an $i_0 \in \{1, \dots, n\}$ such that $a(\pi_{i_0,p}) < v_p(N)$.

Let Φ_{new} be the automorphic form corresponding to the global holomorphic, paramodular newform in π_{i_0}. De-adelizing Φ_{new}, we obtain a Siegel modular form f_{new} with the desired properties. □

We have now proven that Table 4 gives true dimensions and not just upper bounds. Once we know that the dimension of $S_k(K(N))^\epsilon$ agrees with our upper bound, we have $\mathcal{J}^\epsilon_{u_0}(\mathbb{F}_p) = S_k(K(N))^\epsilon[u_0]_p$ and can use the improved u_1^ϵ in Table 3 for which the projection $\text{proj}^{u_0}_{u_1^\epsilon} : \mathcal{J}^\epsilon_{u_0}(\mathbb{F}_p) \to \mathcal{J}^\epsilon_{u_1^\epsilon}(\mathbb{F}_p)$ injects. It follows that $\text{proj}^{\infty}_{u_1^\epsilon} : S_k(K(N))^\epsilon \to S_k(K(N))^\epsilon[u_1^\epsilon]$ injects. With these improved u_1^ϵ, we run Jacobi restriction over $\mathbb{Q}$ to $u = 3u_1^\epsilon$ Jacobi coefficients and break $S_k(K(N))^\epsilon$ into $T(3)$-eigenspaces by verifying the hypotheses of Proposition 3.6.1 and using $\hat{T}(3)$. We stress that we postpone running Jacobi restriction over $\mathbb{Q}$ until we have the improved u_1^ϵ from Table 3 available for $S_k(K(N))^\epsilon$. We are eventually forced to run Jacobi restriction over $\mathbb{Q}$ however, in order to compute Hecke eigenspaces. Once we have $S_k(K(N))^\epsilon$ broken into one-dimensional eigenspaces, we can revert, if we wish, to using $\mathcal{T}(q)$ to compute further good rational eigenvalues inside $\mathcal{J}^\epsilon_{qu_1^\epsilon}(\mathbb{F}_p)$. The point here is that, for $T(q)f = \lambda_q f$, good eigenvalues have simple archimedean bounds $|\lambda_q| \le (1+q)(1+q^2)$ (see [Freitag 1983], page 269, Hilfsatz 4.8), and $q^{k-3}\lambda_q$ is integral for $k \ge 2$. In the next section, however, we are more interested in computing eigenvalues at the bad primes, as a step toward identifying the local representations.

3.11. ***Nonlift newforms.*** From Table 4, we can count how many of each dimension of nonlifts are oldforms from lower levels using the global theory of newforms in [Roberts and Schmidt 2006]. Table 5 breaks $S_k(K(16))^\pm$ into the dimension of newforms and oldforms.

By computing the eigenvalue λ_3 for all the nonlift eigenforms, we are able to distinguish the newforms

	$K(16)^+$		$K(16)^-$	
k	new	old	new	old
6	1	0	0	0
7	0	0	2	0
8	5	0	0	0
9	0	1	7	1
10	11	2	0	2
11	1	3	14	5
12	20	7	1	5
13	3	7	25	9
14	32	14	4	10

Table 5. Breakdown into new and old *nonlift* eigenforms for $S_k(K(16))^\pm$.

	$K(4)$		$K(8)$	
k	$+$	$-$	$+$	$-$
9				-2760
10			-18360 -3672	
11		-13464		$-24(781 \pm 128\sqrt{55})$
12	-88488		-14760 -229032 $-504(-65 \pm 64\sqrt{6})$	
13		-154440	-685224	-271944 $\alpha_{13,8}$ (degree 4)
14	-1422360 -319896		-1176984 199368 $216(1231 \pm 8\sqrt{1129})$ $\alpha_{14,8}$ (degree 3)	-1580472

Table 6. Eigenvalues $3^{k-3}\lambda_3$ of nonlift newforms. Here $\alpha_{13,8}$ represents the four roots of $151059326544225331200 - 28599118413428736x - 271045699200x^2 + 463392x^3 + x^4$ and $\alpha_{14,8}$ represents the three roots of $70155550286581248 - 1194997748544x + 186408x^2 + x^3$.

from the oldforms. See Table 6 for the eigenvalues of nonlift newforms for $S_k(K(4))$ and $S_k(K(8))$ for $k \leq 14$. Note that there are no nonlifts for $S_k(K(N))$ for $N \in \{1, 2\}$ and $k \leq 14$. The eigenvalues of the nonlift newforms for $S_k(K(16))$ with $k \leq 14$ are in Table 7 along with other eigenvalues. We were able to easily distinguish the newforms because it turns out that these newforms have different λ_3 eigenvalues than the oldforms of the same level.

3.12. ***Computing $T_{0,1}$ and $T_{1,0}$.*** The global Hecke operators at the bad primes have their origin in the local theory [Roberts and Schmidt 2006]. The global operators $T_{0,1}(p)$ and $T_{1,0}(p)$ at a bad prime p were defined and studied in [PSY 2018], where eigenvalues were computed that required information from Fourier expansions at multiple zero-dimensional cusps. From Proposition 5.2 of [PSY 2018], the two bad Hecke operators $T_{0,1}(2)$ and $T_{1,0}(2)$ may be written on $S_k(K(16))$ as

$$T_{0,1}F = \sum_{x,y,z\in\{0,1\}} F \mid \begin{bmatrix} 1 & 0 & x & y \\ 0 & 1 & y & z/16 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} + \sum_{x,z\in\{0,1\}} F \mid \begin{bmatrix} 2 & 0 & 0 & 0 \\ x & 1 & 0 & z/16 \\ 0 & 0 & 1 & -x \\ 0 & 0 & 0 & 2 \end{bmatrix} + \sum_{x,y\in\{0,1\}} F \mid \begin{bmatrix} 1 & -16y & x & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 16y & 1 \end{bmatrix}$$
$$+ F \mid \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} + F \mid \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 8 & 16 & 7 & -3 \\ -3 & -8 & 1 & -5/16 \\ 0 & 0 & 1 & -3/8 \\ 0 & 0 & 2 & -1 \end{bmatrix},$$

$$T_{1,0}F=\sum_{\substack{x,y\in\{0,1\}\\ z\in\{0,1,2,3\}}} F\,\Bigg|\begin{bmatrix}2&0&0&2y\\x&1&y&-xy+z/16\\0&0&2&-2x\\0&0&0&4\end{bmatrix}+\sum_{x,y\in\{0,1\}} F\,\Bigg|\begin{bmatrix}1&-16y&0&0\\-x/2&1+8xy&y/2&1/32\\0&0&1+8xy&x/2\\0&0&16y&1\end{bmatrix}$$
$$+\sum_{y\in\{0,1\}} F\,\Bigg|\begin{bmatrix}2&-32y&0&0\\0&4&0&0\\0&0&2&0\\0&0&16y&1\end{bmatrix}+F\,\Bigg|\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&2&1&0\\2&0&0&1\end{bmatrix}\begin{bmatrix}8&32&14&-3\\-3&-16&2&-5/16\\0&0&2&-3/8\\0&0&4&-1\end{bmatrix}+F\,\Bigg|\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&2&1&0\\2&0&0&1\end{bmatrix}\begin{bmatrix}8&32&22&-3\\-3&-16&-1&-5/16\\0&0&2&-3/8\\0&0&4&-1\end{bmatrix}.$$

The zero-dimensional cusps of $K(16)$ are given by the disjoint union

$$\mathrm{GSp}(4,\mathbb{Q})^+=K(16)\,GP_{2,0}(\mathbb{Q})\cup K(16)C_0(2)GP_{2,0}(\mathbb{Q})\cup K(16)C_0(4)GP_{2,0}(\mathbb{Q})$$

(see [Poor and Yuen 2013], Theorem 1.3), where

$$C_0(m)=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&m&1&0\\m&0&0&1\end{bmatrix};\quad GP_{2,0}(R)=\begin{bmatrix}*&0&*&*\\ *&*&*&*\\ *&0&*&*\\0&0&0&*\end{bmatrix}\cap\mathrm{GSp}(4,R).$$

The difficulty in computing $T_{0,1}F$ and $T_{1,0}F$ is that although most of the coset representatives defining $T_{0,1}$ and $T_{1,0}$ lie in the first cusp, a few lie in the second. As in [PSY 2018], we overcome this difficulty by using the technique of restriction to a modular curve to compute the restrictions $F(s\tau+s')$ and $(T_{0,1}F)(s\tau+s')$ for some serviceable choice of s, s'. The point is that it is straightforward to compute $(F\,|\,u)(s\tau+s')$ when $u\in GP_{2,0}(\mathbb{Q})$, but a trick is required to compute $(F\,|\,C_0(2)u)(s\tau+s')$ for the last coset representative in $T_{0,1}$. The strategy of Section 4.2 in [PSY 2018] is to access the cusp $K(N)C_0(m)GP_{2,0}(\mathbb{Q})$ by finding $\sigma=\left[\begin{smallmatrix}\alpha&\beta\\ \gamma&\delta\end{smallmatrix}\right]\in\mathrm{SL}(2,\mathbb{Z})$ and a positive definite $s_0\in\left[\begin{smallmatrix}\mathbb{Z}&\mathbb{Z}\\ \mathbb{Z}&\frac{1}{N}\mathbb{Z}\end{smallmatrix}\right]$ such that $\left[\begin{smallmatrix}\alpha I&\beta s_0\\ \gamma s_0^{-1}&\delta I\end{smallmatrix}\right]\in K(N)C_0(m)W_0$ for some $W_0\in GP_{2,0}(\mathbb{Q})$. Setting

$$W_1=\left[\begin{smallmatrix}A_1&B_1\\0&D_1\end{smallmatrix}\right]=u^{-1}W_0\quad\text{and}\quad s\tau+s'=W_1\langle s_0\tau\rangle=(A_1s_0\tau+B_1)D_1^{-1},$$

it formally follows that

$$(F|_kC_0(m)u)\,(s\tau+s')=\det(A_1D_1)^{-k/2}\det(D_1)^k\,(g|_{2k}\sigma)(\tau),$$

for $g(\tau)=F(s_0\tau)$. For ℓ with $\ell s_0^{-1}\in\left[\begin{smallmatrix}\mathbb{Z}&N\mathbb{Z}\\ N\mathbb{Z}&N\mathbb{Z}\end{smallmatrix}\right]$, we have $g\in S_{2k}\,(\Gamma_0(\ell))$, and we have reduced the problem of specializing F at the $C_0(m)$-cusp to transforming an elliptic modular form.

By choosing $\ell=16$ and σ, s_0, W_0, s, s' as

$$\sigma=\left[\begin{smallmatrix}3&1\\8&3\end{smallmatrix}\right],\ s_0=\left[\begin{smallmatrix}4&1\\1&1/2\end{smallmatrix}\right],\ W_0=\begin{bmatrix}-8&8&-1&6\\1/2&0&-2&-33/16\\0&0&0&1/8\\0&0&2&2\end{bmatrix},\ s=\begin{bmatrix}58&-41/2\\-41/2&29/4\end{bmatrix},\ s'=\begin{bmatrix}41/2&-29/4\\-29/4&81/32\end{bmatrix},$$

we get that

$$\left(F\,|_k\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&2&1&0\\2&0&0&1\end{bmatrix}\begin{bmatrix}8&16&7&-3\\-3&-8&1&-5/16\\0&0&1&-3/8\\0&0&2&-1\end{bmatrix}\right)(s\tau+s')=\left(\tfrac{1}{4}\right)^{-k/2}(1)^k(g|_{2k}\sigma)(\tau),$$

where $g(\tau)=F(s_0\tau)\in S_{2k}(\Gamma_0(16))$. We therefore need to be able to work with cusp forms in $S_{2k}(\Gamma_0(16))$, namely we need to compute a basis of $S_{2k}(\Gamma_0(16))$ and the action of σ on this basis. We show how to do this in Lemma 3.12.1.

To be able to compute the restrictions $F(s\tau+s')$ and $(T_{1,0}F)(s\tau+s')$, for $F\in S_k(K(16))$ and some choice of s, s', we follow the instructions of Section 4.4 in [PSY 2018]. For $T_{1,0}$, the delicate issue is simultaneously computing $(F\,|\,C_0(2)u)(s\tau+s')$ for the last two coset representatives in $T_{1,0}$. By choosing $\ell=16$ and σ, s_0, s, s', τ_0, W_0 as

$$\sigma=\left[\begin{smallmatrix}3&1\\8&3\end{smallmatrix}\right],\;\; s_0=\left[\begin{smallmatrix}10&3\\3&1\end{smallmatrix}\right],\;\; \tau_0=1/2,$$

$$s=\begin{bmatrix}9441370&-2347216\\-2347216&4668325/8\end{bmatrix},\;\; s'=\begin{bmatrix}3152523&-3134991/4\\-3134991/4&12470225/64\end{bmatrix},\;\; W_0=\begin{bmatrix}-24&8&-65&0\\-1055/2&176&-1739&-14897/16\\0&0&-44&-1055/8\\0&0&2&6\end{bmatrix},$$

we get the following, for $g(\tau)=F(s_0\tau)\in S_{2k}\,(\Gamma_0(16))$,

$$\left(F\,|_k\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&2&1&0\\2&0&0&1\end{bmatrix}\begin{bmatrix}8&32&14&-3\\-3&-16&2&-5/16\\0&0&2&-3/8\\0&0&4&-1\end{bmatrix}\right)(s\tau+s')=\left(\tfrac14\right)^{-k/2}(1)^k(g|_{2k}\sigma)(\tau),$$

$$\left(F\,|_k\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&2&1&0\\2&0&0&1\end{bmatrix}\begin{bmatrix}8&32&22&-3\\-3&-16&-1&-5/16\\0&0&2&-3/8\\0&0&4&-1\end{bmatrix}\right)(s\tau+s')=\left(\tfrac14\right)^{-k/2}(1)^k(g|_{2k}\sigma)(\tau+\tau_0).$$

The last thing we need before using this choice to compute $T_{0,1}F$ is a knowledge of how forms in $M_k(\Gamma_0(16))$ transform by $\sigma=\left[\begin{smallmatrix}3&1\\8&3\end{smallmatrix}\right]$. We discuss the ring generators of $M(\Gamma_0(16))=\bigoplus_{k=0}^{\infty}M_k(\Gamma_0(16))$. Let

$$E_2(\tau)=1-24\sum_{n=1}^{\infty}\sigma(n)q^n=1-24q-72q^2-96q^3-168q^4-144q^5-\cdots$$

be the nearly modular weight two Eisenstein series transforming, for all $\left[\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right]\in\mathrm{SL}(2,\mathbb{Z})$, by

$$\left(E_2|_2\left[\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right]\right)(\tau)=E_2(\tau)-\frac{3}{\pi^2}\left(\frac{2\pi ic}{c\tau+d}\right). \tag{36}$$

For $d>1$, we define $E^-_{2,d}\in M_2(\Gamma_0(d))$ by $E^-_{2,d}(\tau)=\frac{1}{1-d}\,(E_2(\tau)-dE_2(d\tau))$. We define five elements in $M_2(\Gamma_0(16))$ by

$$\begin{aligned}
a(\tau)&=\tfrac12E^-_{2,2}(\tau)-3E^-_{2,4}(\tau)+\tfrac72E^-_{2,8}(\tau)=1-24q^2+24q^4-96q^6+24q^8-144q^{10}+\cdots\\
b(\tau)&=-\tfrac{1}{48}E^-_{2,2}(\tau)+\tfrac{7}{48}E^-_{2,8}(\tau)-\tfrac58E^-_{2,16}(\tau)+\tfrac12\vartheta\left[\begin{smallmatrix}2&0&0&0\\0&2&0&0\\0&0&8&0\\0&0&0&8\end{smallmatrix}\right](\tau)=q-4q^3+6q^5-8q^7+13q^9\cdots\\
c(\tau)&=-\tfrac16E^-_{2,2}(\tau)+\tfrac76E^-_{2,8}(\tau)=1+8q^2+24q^4+32q^6+24q^8+48q^{10}+\cdots\\
d(\tau)&=\tfrac{1}{16}E^-_{2,2}(\tau)-\tfrac{1}{16}E^-_{2,4}(\tau)=q+4q^3+6q^5+8q^7+13q^9+\cdots\\
e(\tau)&=\tfrac14E^-_{2,4}(\tau)-\tfrac74E^-_{2,8}(\tau)+\tfrac52E^-_{2,16}(\tau)=1-8q^4+24q^8-328q^{12}+\cdots.
\end{aligned}$$

The theta series $\vartheta[Q]$ of an even m-by-m quadratic form, used above to define basis element b, is defined by $\vartheta[Q](\tau)=\sum_{n\in\mathbb{Z}^m}e\left(\frac12Q[n]\tau\right)$. If ℓQ^{-1} is also even then $\vartheta[Q]\in M_{m/2}(\Gamma_0(\ell),\chi)$ for some character χ. The character is trivial when $\det(Q)$ is a square and $4\mid m$, see [Freitag 1983], page 203. Using Satz 0.3 of [Freitag 1983], we also have, for even m,

$$\vartheta[Q]|F_\ell=\ell^{m/4}\det(Q)^{-1/2}(-i)^{m/2}\vartheta[\ell Q^{-1}],\qquad\text{for } F_\ell=\tfrac{1}{\sqrt{\ell}}\left[\begin{smallmatrix}0&-1\\\ell&0\end{smallmatrix}\right]. \tag{37}$$

A D_4-subgroup of the normalizer of $\Gamma_0(16)$ in $\mathrm{SL}(2,\mathbb{Q})$, modulo $\langle \pm I, \Gamma_0(16)\rangle$, acts on $M_k(\Gamma_0(16))$. This representation of D_4 on $M_2(\Gamma_0(16))$ is 5-dimensional and decomposes into a 2-dimensional irreducible representation and three 1-dimensional representations. The basis of $M_2(\Gamma_0(16))$ defined above was selected to decompose this representation into its irreducible components.

Lemma 3.12.1. *The graded ring $M(\Gamma_0(16))$ consists of homogeneous polynomials in the five elements $a, b, c, d, e \in M_2(\Gamma_0(16))$, subject to the six relations*

$$2e^2 = c^2 + ac, \quad 32d^2 = c^2 - ac, \quad c^2 = a^2 + 64b^2, \quad cd = 2be - ad, \quad ce = ae + 32bd, \quad de = bc.$$

Every element in $M_k(\Gamma_0(16))$ can be uniquely written as

$$P_k(a,b) + C_{k-2}(a,b)c + D_{k-2}(a,b)d + E_{k-2}(a,b)e,$$

where P_k is a homogeneous polynomial of degree $k/2$ and the C_{k-2}, D_{k-2}, E_{k-2} are homogeneous of degree $(k-2)/2$. The Fricke involution $F = \left[\begin{smallmatrix} 0 & -1/4 \\ 4 & 0 \end{smallmatrix}\right]$ and the translation $A = \left[\begin{smallmatrix} 1 & 1/2 \\ 0 & 1 \end{smallmatrix}\right]$ normalize $\Gamma_0(16)$ and generate a subgroup isomorphic to the dihedral group D_4, with $T = AF = \left[\begin{smallmatrix} 2 & -1/4 \\ 4 & 0 \end{smallmatrix}\right]$ of order four, and $\sigma = T^3F = \left[\begin{smallmatrix} 3 & 1 \\ 8 & 3 \end{smallmatrix}\right]$ of order two. For the representation $\rho : D_4 \to \mathrm{GL}(5,\mathbb{C})$ defined by $(a,b,c,d,e)|_2 g = (a,b,c,d,e)\rho(g)$, we have

$$\rho(A) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \rho(F) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -4 \\ 0 & 0 & 0 & -1/4 & 0 \end{bmatrix}, \quad \rho(T) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & -1/4 & 0 \end{bmatrix}, \quad \rho(\sigma) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

Proof. The transformation under A is obvious and the transformation under F may be worked out using (36) and (37). A helpful intermediate step is $(E^-_{2,d}|F)(\tau) = -\frac{16}{d}E^-_{2,d}\big(\frac{16}{d}\tau\big)$. The normalizer in $\mathrm{SL}(2,\mathbb{Q})$ of $\Gamma_0(16)$, modulo $\langle \pm I, \Gamma_0(16)\rangle$, contains a dihedral group of order 8: $\langle A, F\rangle = \langle T, \sigma\rangle$. The index of $\Gamma_0(16)$ in $\mathrm{SL}(2,\mathbb{Z})$ is 24, so, by the Valence Inequality, to prove equality in $M_k(\Gamma_0(16))$ it suffices to check the equality of the first $2k+1$ Fourier coefficients. In this way we verify the six given relations and the images of ρ.

Every modular form in $M_k(\Gamma_0(16))$ that can be written as a polynomial in a, b, c, d, e, may be written in the form $P_k(a,b) + C_{k-2}(a,b)c + D_{k-2}(a,b)d + E_{k-2}(a,b)e$, by appying the given relations in the order given. We will show that no nontrivial relation of the given form can be zero. First, by applying T^2, we would have both $P_k(a,b) + C_{k-2}(a,b)c = 0$ and $D_{k-2}(a,b)d + E_{k-2}(a,b)e = 0$. Second, applying T to the first we obtain $P_k(a,b) - C_{k-2}(a,b)c = 0$ and hence $P_k(a,b) = C_{k-2}(a,b) = 0$. The modular forms a and b have the same weight, and so are algebraically independent because b/a is nonconstant. Hence the polynomials P_k and C_{k-2} are also trivial. Third, applying T to the second we obtain $D_{k-2}(a,b)(4e) - E_{k-2}(a,b)(d/4) = 0$ as well. Over the field of meromorphic functions, we thus have $E_{k-2}(a,b) = \pm 4D_{k-2}(a,b)$ and this is also an equality among holomorphic functions. From $0 = D_{k-2}(a,b)d + E_{k-2}(a,b)e = D_{k-2}(a,b)(d \pm 4e)$, we conclude that D_{k-2} and E_{k-2} are zero as polynomials. The dimension of $\mathbb{C}[a,b,c,d,e] \cap M_k(\Gamma_0(16))$ is then $\big(\frac{k}{2}+1\big) + 3\big(\frac{k-2}{2}+1\big) = 2k+1$. By the Riemann–Roch theorem, $\dim M_k(\Gamma_0(16)) = 2k+1$ for even $k \geq 0$, and thus $M(\Gamma_0(16)) = \mathbb{C}[a,b,c,d,e]$ as graded rings. □

We have all the ingredients to apply the techniques of Section 4.2 and 4.4 of [PSY 2018] to compute the eigenvalues $\lambda_{0,1}$ and $\lambda_{1,0}$. We successfully computed the eigenvalues $\lambda_{0,1}$ and $\lambda_{1,0}$ of the nonlift newforms in $S_k(K(16))^{\pm}$ for $k \leq 14$. The results are in Table 7. By applying the knowledge of these

k	AL	$3^{k-3}\,\lambda_3$	$\lambda_{0,1}$	$\lambda_{1,0}$	type
6	$+$	-96	-5	0	X
7	$-$	-600	-2	-4	XIa
	$-$	-144	-3	0	I, IIa, or X
8	$+$	-1992	0	-4	VII, VIIIa or IXa
	$+$	912	-3	0	X
	$+$	-168	-2	-4	XIa
	$+$	$-864 \pm 112\sqrt{33}$	$1/8(-7 \mp \sqrt{33})$	0	X
9	$-$	-8136	2	-4	XIa
	$-$	5856	-5	0	I, IIa, or X
	$-$	-2280	0	-4	sc(16)
	$-$	-1920	$1/4$	0	I, IIa, or X
	$-$	1464	-2	-4	XIa
	$-$	$\pm 480\sqrt{33}$	$1/4(-3 \mp \sqrt{33})$	0	I, IIa, or X
10	$+$	-12888	2	-4	XIa
	$+$	5928	-2	-4	XIa
	$+$	-3768	0	-4	VII, VIIIa or IXa
	$+$	-1080	0	-4	VII, VIIIa or IXa
	$+$	$7248 \pm 240\sqrt{505}$	$1/8(-19 \pm \sqrt{505})$	0	X
	$+$	$\alpha_{10,16}$ (degree 5)	t_{10}	0	X
11	$+$	-66096	$-29/8$	0	X
	$-$	8040	0	-4	sc(16)
	$-$	$24(-1245 \pm 32\sqrt{21})$	2	-4	XIa
	$-$	$120(111 \pm 8\sqrt{69})$	-2	-4	XIa
	$-$	-73584	$9/2$	0	I, IIa, or X
	$-$	18768	1	0	I, IIa, or X
	$-$	35568	$-3/4$	0	I, IIa, or X
	$-$	$48(425 \pm 2\sqrt{3961})$	$1/32(-107 \pm \sqrt{3961})$	0	I, IIa, or X
	$-$	$\alpha_{11,16}$ (degree 4)	t_{11}	0	I, IIa, or X

Table 7. Eigenvalues λ_3, $\lambda_{0,1}$ and $\lambda_{1,0}$ of nonlift newforms in $S_k(K(16))^{\pm}$. The algebraic numbers $\alpha_{10,16}$, $\alpha_{11,16}$ and the corresponding eigenvalues t_{10}, t_{11} are given below. (Table continues on the next page.)

Symbolic constants in Table 7, for $k = 10$, 11: minimal polynomial of α_* and eigenvalues	
$\alpha_{10,16}$:	$-392100597530099712 + 36717761396736000x - 1936322592768x^2 - 384208896x^3 + 12000x^4 + x^5$
$t_{10} =$	$(200684470423235227287552 + 94255611784369274880\alpha + 2115778851231744\alpha^2 - 1410266234784\alpha^3 - 54792385\alpha^4)/410907531887271468859392$
$\alpha_{11,16}$:	$332724999250575360 - 1154234880x^2 + x^4$
$t_{11} =$	$(858199620022272 + 28477875456\alpha - 1490544\alpha^2 - 53\alpha^3)/21539386294272$

k	AL	$3^{k-3}\lambda_3$	$\lambda_{0,1}$	$\lambda_{1,0}$	type
12	$+$	-12456	0	-4	VII, VIIIa or IXa
	$+$	$72(819 \pm 64\sqrt{85})$	0	-4	VII, VIIIa or IXa
	$+$	$72(-521 \pm 128\sqrt{5})$	2	-4	XIa
	$+$	$72(831 \pm 8\sqrt{85})$	-2	-4	XIa
	$+$	$\alpha_{12,16,a}$ (degree 5)	$t_{12,a}$	0	X
	$+$	$\alpha_{12,16,b}$ (degree 8)	$t_{12,b}$	0	X
	$-$	-185616	$-21/8$	0	I, IIa, or X
13	$+$	-183168	$-33/8$	0	X
	$+$	$-144(3879 \pm 41\sqrt{609})$	$(-53 \pm \sqrt{609})/32$	0	X
	$-$	-220968	2	-4	XIa
	$-$	$72(-333 \pm 80\sqrt{609})$	2	-4	XIa
	$-$	$\alpha_{13,16,a}$ (degree 3)	-2	-4	XIa
	$-$	$\alpha_{13,16,b}$ (degree 3)	0	-4	sc(16)
	$-$	0	$3/2$	0	I, IIa, or X
	$-$	725184	-1	0	I, IIa, or X
	$-$	$\alpha_{13,16,c}$ (degree 6)	$t_{13,c}$	0	I, IIa, or X
	$-$	$\alpha_{13,16,d}$ (degree 4)	$t_{13,d}$	0	I, IIa, or X
	$-$	$\alpha_{13,16,e}$ (degree 4)	$t_{13,e}$	0	I, IIa, or X
14	$+$	517320	2	-4	XIa
	$+$	527688	-2	-4	XIa
	$+$	$216(-597 \pm 16\sqrt{51})$	2	-4	XIa
	$+$	$24(40387 \pm 320\sqrt{25561})$	-2	-4	XIa
	$+$	-499608	0	-4	VII, VIIIa or IXa
	$+$	$216(2927 \pm 56\sqrt{3889})$	0	-4	VII, VIIIa or IXa
	$+$	$24(20759 \pm 88\sqrt{8689})$	0	-4	VII, VIIIa or IXa
	$+$	$\alpha_{14,16,a}$ (degree 8)	$t_{14,a}$	0	X
	$+$	$\alpha_{14,16,b}$ (degree 13)	$t_{14,b}$	0	X
	$-$	-2434968	0	-4	sc(16)
	$-$	-927072	$-17/8$	0	I, IIa, or X
	$-$	$-432(1935 \pm 23\sqrt{2377})$	$(-97 \pm \sqrt{2377})/32$	0	I, IIa, or X

(For the minimal polynomials of the algebraic numbers α_* and the corresponding eigenvalues see [Yuen 2018].)

Table 7, continued.

eigenvalues to Table A.14 of [Roberts and Schmidt 2007], we also identify the possibilities for the corresponding local representations at $p = 2$ of the underlying automorphic representations. Further information on the entries of these tables may be found at [Yuen 2018].

3.13. ***Supercuspidal forms found.*** From Table 7, we see that we found supercuspidal forms in weights 9, 11, 13, 14. The website [Yuen 2018] gives formulas for these supercuspidal forms. For the odd weights

$k = 9, 11, 13$, the supercuspidal form is given as a linear combination of Gritsenko lifts and repeated $T(3)$ images of one or more Borcherds products. For the even weight $k = 14$, the supercuspidal form is given as a linear combination of the repeated $T(3)$ images of one Borcherds product. We also give the formula for the weight 14 supercuspidal form here to provide a bridge to the database [Yuen 2018] and to aid any future reproduction of our results. Let Δ be the cusp form in $S_{12}(\mathrm{SL}_2(\mathbb{Z}))$ normalized to have leading term q. Theta blocks are the invention of Gritsenko, Skoruppa, and Zagier, and the special case we use here may be defined, for $d_j \in \mathbb{N}$, by

$$\mathrm{TB}_k(d_1, d_2, \dots, d_\ell)(\tau, z) = \eta(\tau)^{2k-\ell} \prod_{j=1}^{\ell} \vartheta(\tau, d_j z),$$

where η is the Dedekind eta function and $\vartheta(\tau, z) = \sum_{n \in \mathbb{Z}} (-1)^n q^{(n+1/2)^2/2} \zeta^{n+1/2}$ is the odd Jacobi theta function. A basis $\mathbf{B}$ of $J_{12,16}^{\mathrm{cusp}}$ is given in Table 8 in terms of W_2 and W_3 images of theta blocks.

$\mathrm{TB}_{12}(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 3, 4)\vert W_2$
$\mathrm{TB}_{12}(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3)\vert W_2$
$\mathrm{TB}_{12}(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 8)\vert W_3$
$\mathrm{TB}_{12}(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 4, 7)\vert W_3$
$\mathrm{TB}_{12}(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 4, 4, 4, 5)\vert W_3$
$\mathrm{TB}_{12}(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 3, 3, 3, 4, 4, 4)\vert W_3$
$\mathrm{TB}_{12}(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 7)\vert W_3$
$\mathrm{TB}_{12}(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 5, 5)\vert W_3$
$\mathrm{TB}_{12}(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 3, 3, 3, 4)\vert W_3$
$\mathrm{TB}_{12}(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3)\vert W_3$
$\mathrm{TB}_{12}(1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 5)\vert W_3$
$\mathrm{TB}_{12}(1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3)\vert W_3$

Table 8. A basis $\mathbf{B}$ of $J_{12,16}^{\mathrm{cusp}}$.

Define the weight-zero weakly holomorphic form $\psi_{14} \in J_{0,16}^{\mathrm{wh}}(\mathbb{Z})$ by

$$\phi_{14} = \mathrm{TB}_{14}(1, 1, 1, 1, 1, 1, 2, 2, 3, 3); \qquad \psi_{14} = \frac{\phi_{14} | V_2}{\phi_{14}} + \frac{\mathbf{b}_{14} \cdot \mathbf{B}}{\Delta},$$

where the vector $\mathbf{b}_{14}$ is given by

$$\mathbf{b}_{14} = \frac{1}{279268001096663167080660} \cdot (-1155865602408281 7198192, -1056598136932746 2562477, -2926740930944006282896,$$
$$9167023003084404792024, 9262973271453152666448, 5762211536895867593392,$$
$$2926740930944006282896, -5759260672816406 31444, 1918503995959964699328,$$
$$-1300786643303689 05144, 1584969973757747 48880, 3516352762722050 84768)$$

We have $\mathrm{Borch}(\psi_{14}) \in S_{14}(K(16))^-$. It happens that $\{T(3)^j \,\mathrm{Borch}(\psi_{14}) : j = 0, \dots, 13\}$ is a basis of the space $S_k(K(16))^-$. We state on the next page the linear combination vector $\mathbf{c}_{14}$ that defines

$$f_{14} = \sum_{j=0}^{13} (\mathbf{c}_{14})_j \, T(3)^j \,\mathrm{Borch}(\psi_{14}) :$$

$$\begin{aligned}\mathbf{c}_{14} = \frac{1}{3725738216385042356336483182434882958372211606631219200\!0} \\ &(3482441572973122342461994879166366309741352437415936 00, \\ &-1231050015269758711257977743259890444383052096717455360, \\ &18757558102267391393392499778171686231151225571010150 4, \\ &1254983988315996708233967338189308356957980874856464384, \\ &-1274790786621908526577376789255161461974879582929551 36, \\ &-4875516113922102296958025126823830268826527115415388 16, \\ &-5141021228456145989487013649851790987626368922445414 4, \\ &6375351289634317218668183191220580480064617628670361 6, \\ &2045229986855668665249903450571345856571047582485708 8, \\ &112323978289166189088890862245481898303267566214348 8, \\ &-429183614695895171861584434488686219409693487083520, \\ &-865110231933857931075636730023128902729602127075 20, \\ &-635577289399001689023352277573478866283690349339 2, \\ &-171792506910670443678820376588540424234035840667)\end{aligned}$$

The relevant definitions for other weights are at the website [Yuen 2018].

We stopped at $k = 14$ because we found a supercuspidal paramodular form in an even weight space of the lowest possible level. Also, weight $k = 14$ for $K(16)$ is on the edge of tractability for the method of Jacobi restriction.

References

[Berger and Klosin 2017] T. Berger and K. Klosin, "Deformations of Saito–Kurokawa type and the paramodular conjecture", preprint, 2017. arXiv

[Borcherds 1998] R. E. Borcherds, "Automorphic forms with singularities on Grassmannians", *Invent. Math.* **132**:3 (1998), 491–562. MR Zbl

[Breeding et al. 2016] J. Breeding, II, C. Poor, and D. S. Yuen, "Computations of spaces of paramodular forms of general level", *J. Korean Math. Soc.* **53**:3 (2016), 645–689. MR Zbl

[Carter 1985] R. W. Carter, *Finite groups of Lie type*, Wiley, New York, 1985. MR Zbl

[DeBacker and Reeder 2009] S. DeBacker and M. Reeder, "Depth-zero supercuspidal L-packets and their stability", *Ann. of Math.* (2) **169**:3 (2009), 795–901. MR Zbl

[Deligne and Lusztig 1976] P. Deligne and G. Lusztig, "Representations of reductive groups over finite fields", *Ann. of Math.* (2) **103**:1 (1976), 103–161. MR Zbl

[Eichler and Zagier 1985] M. Eichler and D. Zagier, *The theory of Jacobi forms*, Progr. Math. **55**, Birkhäuser, Boston, 1985. MR Zbl

[Enomoto 1972] H. Enomoto, "The characters of the finite symplectic group $\mathrm{Sp}(4, q)$, $q = 2^f$", *Osaka Math. J.* **9** (1972), 75–94. MR Zbl

[Freitag 1983] E. Freitag, *Siegelsche Modulfunktionen*, Grundlehren der Math. Wissenschaften **254**, Springer, 1983. MR Zbl

[Gan and Takeda 2011] W. T. Gan and S. Takeda, "The local Langlands conjecture for GSp(4)", *Ann. of Math.* (2) **173**:3 (2011), 1841–1882. MR Zbl

[Gan and Tantono 2014] W. T. Gan and W. Tantono, "The local Langlands conjecture for GSp(4), II: The case of inner forms", *Amer. J. Math.* **136**:3 (2014), 761–805. MR Zbl

[Gritsenko 1995] V. Gritsenko, "Arithmetical lifting and its applications", pp. 103–126 in *Number theory* (Paris 1992–1993), edited by S. David, London Math. Soc. Lecture Note Ser. **215**, Cambridge Univ. Press, 1995. MR Zbl

[Gritsenko and Nikulin 1998] V. A. Gritsenko and V. V. Nikulin, "Automorphic forms and Lorentzian Kac–Moody algebras, II", *Int. J. Math.* **9**:2 (1998), 201–275. MR Zbl

[Gross and Prasad 1992] B. H. Gross and D. Prasad, "On the decomposition of a representation of SO_n when restricted to SO_{n-1}", *Canad. J. Math.* **44**:5 (1992), 974–1002. MR Zbl

[Gross and Reeder 2010] B. H. Gross and M. Reeder, "Arithmetic invariants of discrete Langlands parameters", *Duke Math. J.* **154**:3 (2010), 431–508. MR Zbl

[Henniart and Herb 1995] G. Henniart and R. Herb, "Automorphic induction for $\mathrm{GL}(n)$ (over local non-Archimedean fields)", *Duke Math. J.* **78**:1 (1995), 131–192. MR Zbl

[Ibukiyama and Onodera 1997] T. Ibukiyama and F. Onodera, "On the graded ring of modular forms of the Siegel paramodular group of level 2", *Abh. Math. Sem. Univ. Hamburg* **67** (1997), 297–305. MR Zbl

[Ibukiyama et al. 2013] T. Ibukiyama, C. Poor, and D. S. Yuen, "Jacobi forms that characterize paramodular forms", *Abh. Math. Sem. Univ. Hamburg* **83**:1 (2013), 111–128. MR Zbl

[Igusa 1962] J.-i. Igusa, "On Siegel modular forms of genus two", *Amer. J. Math.* **84** (1962), 175–200. MR Zbl

[Knightly and Ragsdale 2014] A. Knightly and C. Ragsdale, "Matrix coefficients of depth-zero supercuspidal representations of GL(2)", *Involve* **7**:5 (2014), 669–690. MR Zbl

[Kohnen and Skoruppa 1989] W. Kohnen and N.-P. Skoruppa, "A certain Dirichlet series attached to Siegel modular forms of degree two", *Invent. Math.* **95**:3 (1989), 541–558. MR Zbl

[Lansky and Raghuram 2003] J. Lansky and A. Raghuram, "On the correspondence of representations between $\mathrm{GL}(n)$ and division algebras", *Proc. Amer. Math. Soc.* **131**:5 (2003), 1641–1648. MR Zbl

[Lust 2013] J. Lust, "Depth-zero supercuspidal L-packets for inner forms of GSp_4", *J. Algebra* **389** (2013), 23–60. MR Zbl

[Moy and Prasad 1994] A. Moy and G. Prasad, "Unrefined minimal K-types for p-adic groups", *Invent. Math.* **116**:1-3 (1994), 393–408. MR Zbl

[Moy and Prasad 1996] A. Moy and G. Prasad, "Jacquet functors and unrefined minimal K-types", *Comment. Math. Helv.* **71**:1 (1996), 98–121. MR Zbl

[Poor and Yuen 2013] C. Poor and D. S. Yuen, "The cusp structure of the paramodular groups for degree two", *J. Korean Math. Soc.* **50**:2 (2013), 445–464. MR Zbl

[Poor et al. 2018] C. Poor, J. Shurman, and D. S. Yuen, "Finding all Borcherds product paramodular cusp forms of a given weight and level", preprint, 2018. arXiv

[PSY 2018] C. Poor, R. Schmidt, and D. S. Yuen, "Paramodular forms of level 8 and weights 10 and 12", *Int. J. Number Theory* **14**:2 (2018), 417–467. MR Zbl

[Roberts 2001] B. Roberts, "Global L-packets for GSp(2) and theta lifts", *Doc. Math.* **6** (2001), 247–314. MR Zbl

[Roberts and Schmidt 2006] B. Roberts and R. Schmidt, "On modular forms for the paramodular groups", pp. 334–364 in *Automorphic forms and zeta functions* (Tokyo, 2004), edited by S. Böcherer et al., World Sci., Hackensack, NJ, 2006. MR Zbl

[Roberts and Schmidt 2007] B. Roberts and R. Schmidt, *Local newforms for* GSp(4), Lecture Notes in Math. **1918**, Springer, 2007. MR Zbl

[Roberts and Schmidt 2016] B. Roberts and R. Schmidt, "Some results on Bessel functionals for GSp(4)", *Doc. Math.* **21** (2016), 467–553. MR Zbl

[Rohrlich 1994] D. E. Rohrlich, "Elliptic curves and the Weil–Deligne group", pp. 125–157 in *Elliptic curves and related topics*, edited by H. Kisilevsky and M. R. Murty, CRM Proc. Lecture Notes **4**, Amer. Math. Soc., Providence, RI, 1994. MR Zbl

[Schmidt 2018] R. Schmidt, "Packet structure and paramodular forms", *Trans. Amer. Math. Soc.* **370**:5 (2018), 3085–3112. MR Zbl

[Shimura 1975] G. Shimura, "On the Fourier coefficients of modular forms of several variables", *Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II* **17** (1975), 261–268. MR Zbl

[Skoruppa and Zagier 1989] N.-P. Skoruppa and D. Zagier, "A trace formula for Jacobi forms", *J. Reine Angew. Math.* **393** (1989), 168–198. MR Zbl

[Tunnell 1978] J. B. Tunnell, "On the local Langlands conjecture for GL(2)", *Invent. Math.* **46**:2 (1978), 179–200. MR Zbl

[Vogan 1993] D. A. Vogan, Jr., "The local Langlands conjecture", pp. 305–379 in *Representation theory of groups and algebras*, edited by J. Adams et al., Contemp. Math. **145**, Amer. Math. Soc., Providence, RI, 1993. MR Zbl

[Wilson 2009] R. A. Wilson, *The finite simple groups*, Graduate Texts in Math. **251**, Springer, 2009. MR Zbl

[Yuen 2018] D. S. Yuen, "Degree 2 Siegel paramodular forms of level 16 and weights up to 14", webpage, 2018, http://www.siegelmodularforms.org/pages/degree2/paramodular-level-16.

Received 27 Oct 2018. Revised 3 May 2019.

CRIS POOR:
poor@fordham.edu
Department of Mathematics, Fordham University, Bronx, NY, United States

RALF SCHMIDT:
ralf.schmidt@unt.edu
Department of Mathematics, University of North Texas, Denton, TX, United States

DAVID S. YUEN:
yuen888@hawaii.edu
Department of Mathematics, University of Hawaii, Honolulu, HI, United States

Moscow Journal of Combinatorics and Number Theory
Vol. 8, No. 4, 2019
dx.doi.org/10.2140/moscow.2019.8.325

Generalized Beatty sequences and complementary triples

Jean-Paul Allouche and F. Michel Dekking

A generalized Beatty sequence is a sequence V defined by $V(n) = p\lfloor n\alpha\rfloor + qn + r$, for $n = 1, 2, \dots$, where α is a real number, and p, q, r are integers. Such sequences occur, for instance, in homomorphic embeddings of Sturmian languages in the integers.

We consider the question of characterizing pairs of integer triples (p, q, r), (s, t, u) such that the two sequences $V(n) = (p\lfloor n\alpha\rfloor + qn + r)$ and $W(n) = (s\lfloor n\alpha\rfloor + tn + u)$ are complementary (their image sets are disjoint and cover the positive integers). Most of our results are for the case that α is the golden mean, but we show how some of them generalize to arbitrary quadratic irrationals.

We also study triples of sequences $V_i = (p_i\lfloor n\alpha\rfloor + q_i n + r_i)$, $i = 1, 2, 3$ that are complementary in the same sense.

1. Introduction

A Beatty sequence is the sequence $A = (A(n))_{n\ge1}$, with $A(n) = \lfloor n\alpha\rfloor$ for $n \ge 1$, where α is a positive real number. What Beatty observed is that when $B = (B(n))_{n\ge1}$ is the sequence defined by $B(n) = \lfloor n\beta\rfloor$, with α and β satisfying

$$\frac{1}{\alpha} + \frac{1}{\beta} = 1, \tag{1}$$

then A and B are *complementary* sequences, that is, the sets $\{A(n) : n \ge 1\}$ and $\{B(n) : n \ge 1\}$ are disjoint and their union is the set of positive integers. In particular if $\alpha = \varphi = \frac{1+\sqrt{5}}{2}$ is the golden ratio, this gives that the sequences $(\lfloor n\varphi\rfloor)_{n\ge1}$ and $(\lfloor n\varphi^2\rfloor)_{n\ge1}$ are complementary.

Carlitz, Scoville and Hoggatt [Carlitz et al. 1972] studied the monoid generated by $A = (A(n))_{n\ge1}$ and $B = (B(n))_{n\ge1}$ for the composition of sequences in the case where α is the golden ratio. (The composition of two integer sequences $U = (U(n))_{n\ge1}$ and $V = (V(n))_{n\ge1}$ is the sequence $UV := U \circ V = (U(V(n)))_{n\ge1}$, so that the monoid generated by A and B is composed of sequences like $A^k B^j A^\ell \dots$, where $A^k = AA\dots A$ is the composition of k copies of A.)

Theorem 1 [Carlitz et al. 1972, Theorem 13, p. 20]. *Let $U = (U(n))_{n\ge1}$ be a composition of copies of $A = (\lfloor n\varphi\rfloor)_{n\ge1}$ and $B = (\lfloor n\varphi^2\rfloor)_{n\ge1}$, containing i occurrences of A and j occurrences of B. Then for all $n \ge 1$*

$$U(n) = F_{i+2j}A(n) + F_{i+2j-1}\,n - \lambda_U,$$

where F_k are the Fibonacci numbers ($F_0 = 0$, $F_1 = 1$, $F_{n+2} = F_{n+1} + F_n$) and λ_U is a constant.

MSC2010: 11B83, 11B85, 11D09, 11J70, 68R15.
Keywords: generalized Beatty sequences, complementary pairs and triples, morphic words, return words, Kimberling transform.

Definition 1. A *generalized Beatty sequence* is any sequence V of the type $V(n) = p(\lfloor n\alpha \rfloor) + qn + r$, $n \geq 1$, where α is a real number and p, q, and r are integers.

Two examples of generalized Beatty sequences are $U = AA$ and $U = AB$, where Theorem 1 gives $AA(n) = A(n) + n - 1$, and $AB(n) = 2A(n) + n$. These two formulas directly imply the following result.

Corollary 2. *Let V be a generalized Beatty sequence given by $V(n) = p(\lfloor n\varphi \rfloor) + qn + r$, $n \geq 1$. Then VA and VB are generalized Beatty sequences with parameters $(p_{VA}, q_{VA}, r_{VA}) = (p+q, p, r-p)$ and $(p_{WA}, q_{WA}, r_{WA}) = (2p+q, p+q, r)$.*

As an extension of Beatty's observation the following natural questions can be asked.

Question A. Let α be an irrational number, and let A defined by $A(n) = \lfloor n\alpha \rfloor$ for $n \geq 1$ be the Beatty sequence of α. Let Id defined by $\mathrm{Id}(n) = n$ be the identity map on the integers. For which pairs of integer triples (p, q, r) (s, t, u) are the two sequences $V = pA + q\,\mathrm{Id} + r$ and $W = sA + t\,\mathrm{Id} + u$ complementary?

Question B. For which triples of integer triples (p_1, q_1, r_1), (p_2, q_2, r_2), (p_3, q_3, r_3) are the sequences

$$V_i = p_i A + q_i\,\mathrm{Id} + r_i,\ i = 1, 2, 3,$$

a complementary triple? That is, when do they determine disjoint sets whose union is the positive integers? (Further, when is this partition "nice", or, in the terminology of [Fraenkel 1994], a *nice integer disjoint covering system*?)

Remark 3. The theorem of Carlitz, Scoville and Hoggatt above was rediscovered by Kimberling [2008, Theorem 5, p. 3], to whom it is attributed in, e.g., [Fraenkel 2010a, p. 575; Fraenkel 2010b, p. 647; Larsson et al. 2015, p. 20–21]. This was corrected in [Ballot 2017, Theorem 2, p. 2]. Theorem 1 in [Griffiths 2015] is also a special case of the same theorem of Carlitz et al.

Remark 4. Different generalizations of Beatty sequences are considered in [Mercer 1978; Artstein-Avidan et al. 2008; Kimberling 2011; Hildebrand et al. 2018].

Remark 5. One can ask whether the monoid generated by other complementary sequences by composition can be written as a subset of the set of linear combinations of a finite number of elements. Some answers for Beatty sequences can be found in a rich paper of Fraenkel [1994] (see, e.g., p. 645). Another, possibly unexpected, example is given by the Thue–Morse sequence. Namely, calling odious (resp. evil) the integers whose binary expansion contains an odd (resp. even) number of 1's, it was proved in [Allouche et al. 2016, Corollaries 1 and 3] that the sequences $(A(n))_{n\geq 0}$ and $(B(n))_{n\geq 0}$ of odious and evil numbers satisfy for all n

$$\begin{aligned} &A(n) = 2n + 1 - t(n), && B(n) = 2n + t(n), && A(n) - B(n) = 1 - 2t(n), \\ &A(A(n)) = 2A(n), && B(B(n)) = 2B(n), && A(B(n)) = 2B(n) + 1, \quad B(A(n)) = 2A(n) + 1, \end{aligned}$$

where $(t(n))_{n\geq 0}$ is the Thue–Morse sequence, i.e., the characteristic function of odious integers. (This sequence can be defined by $t(0) = 0$ and for all $n \geq 0$, $t(2n) = t(n)$ and $t(2n+1) = 1 - t(n)$.) This easily implies that any finite composition of $(A(n))_{n\geq 0}$ and $(B(n))_{n\geq 0}$ can be written as $(\alpha A(n) + \beta B(n) + \gamma)_{n\geq 0}$, since $t(A(n)) = 1$ and $t(B(n)) = 0$ for all n.

2. Complementary pairs

Let α be an irrational number, and let A be the Beatty sequence of α. In this section we consider Question A of the Introduction, which we call the Complementary pair problem.

In what follows we will require that as a function $A : \mathbb{N} \to \mathbb{N}$ be injective, since we then have a one-to-one correspondence between sequences and subsets of $\mathbb{N}$. (See [Kimberling and Stolarsky 2016] for noninjective Beatty sequences.)

In the case that V and W are increasing, we will also require, without loss of generality, that $V(1) = 1$. Solutions (p, q, r, s, t, u) with $p = 0$ or $s = 0$ will be called *trivial*.

The homogeneous Sturmian sequence generated by a real number $\alpha \in (0, 1)$ is the sequence

$$c_\alpha := (\lfloor (n+1)\alpha \rfloor - \lfloor n\alpha \rfloor)_{n\geq 1}.$$

(For more about Sturmian sequences, the reader can consult, e.g., [Lothaire 2002, Chapter 2].) A real number α is called a *Sturm number* if $\alpha \in (0, 1)$ is a quadratic irrational number with algebraic conjugate $\overline{\alpha}$ satisfying $\overline{\alpha} \notin (0, 1)$. Sturm numbers have a property that is useful to recognize their generalized Beatty sequences.

Proposition 6 [Crisp et al. 1993; Allauzen 1998]. *Let α be a Sturm number. Then there exists a morphism σ_α on the alphabet $\{0, 1\}$ such that $\sigma_\alpha(c_\alpha) = c_\alpha$.*

In the following we will consider the variants of σ_α on various other alphabets than $\{0, 1\}$, but will not indicate this in the notation. The following lemma is implied trivially by

$$V = pA + q\,\mathrm{Id} + r \quad\Rightarrow\quad V(n+1) - V(n) = p(A(n+1) - A(n)) + q = p\,c_\alpha(n) + q.$$

Lemma 7. *Let α be a Sturm number. Let $V = (V(n))_{n\geq 1}$ be the generalized Beatty sequence defined by $V(n) = p(\lfloor n\alpha \rfloor) + qn + r$, and let ΔV be the sequence of its first differences. Then ΔV is the fixed point of σ_α on the alphabet $\{q, p+q\}$.*

We remark that it can be shown that the first letters of $\sigma_\alpha(0)$ and $\sigma_\alpha(1)$ are equal (see, e.g., [Dekking 2018b]), so σ_α has a unique fixed point. It is also obvious that this fixed point starts with 0 if $\alpha \in (0, 1/2)$, and with 1 if $\alpha \in (1/2, 1)$. For general α, one replaces α with $\breve{\alpha} = \alpha - \lfloor \alpha \rfloor (= \{\alpha\})$. When α is the golden mean $\varphi = (1+\sqrt{5})/2$, the morphism generating the sequence associated to the Sturm number $\breve{\varphi} = \varphi - 1$ is $0 \mapsto 1$, $1 \mapsto 10$, so one has to exchange 0 and 1 if one wishes to compare ΔV with the classical Fibonacci morphism $0 \mapsto 01$, $1 \mapsto 0$. As a special case of Lemma 7 we therefore obtain one direction of the following lemma.

Lemma 8. *Let $V = (V(n))_{n\geq 1}$ be the generalized Beatty sequence defined by $V(n) = p(\lfloor n\varphi \rfloor) + qn + r$, and let ΔV be the sequence of its first differences. Then ΔV is the Fibonacci word on the alphabet $\{2p+q, p+q\}$. Conversely, if x_{ab} is the Fibonacci word on the alphabet $\{a, b\}$, then any V with $\Delta V = x_{ab}$ is a generalized Beatty sequence $V = ((a-b)\lfloor n\varphi \rfloor) + (2b-a)n + r)$ for some integer r.*

Another observation is that the $q\,\mathrm{Id} + r$ part in a generalized Beatty sequence generates arithmetic sequences. The following lemma, which will be useful in proving Theorem 11 below, shows that in some weak sense the Wythoff part pA of a generalized Beatty sequence is orthogonal to its arithmetic sequence part, provided that $\frac{1}{3} < \{\alpha\} < \frac{2}{3}$, where $\{\alpha\} = \alpha - \lfloor \alpha \rfloor$. We prove this for $\frac{4}{3} < \alpha < \frac{5}{3}$.

Lemma 9. *Let α satisfy $\frac{4}{3} < \alpha < \frac{5}{3}$, and let $V = (V(n))_{n\geq 1}$ be the generalized Beatty sequence defined by $V(n) = p(\lfloor n\alpha \rfloor) + qn + r$ with $p \neq 0$, then neither $(V(1), V(2), V(3))$, nor $(V(2), V(3), V(4))$ can be an arithmetic sequence of length* 3.

Proof. When $\frac{3}{2} < \alpha < \frac{5}{3}$, we have $\lfloor \alpha \rfloor = 1$, $\lfloor 2\alpha \rfloor = 3$, and $\lfloor 3\alpha \rfloor = 4$, $\lfloor 4\alpha \rfloor = 6$, so

$$\begin{aligned} V(2) - V(1) &= p\lfloor 2\alpha \rfloor + 2q + r - p\lfloor \alpha \rfloor - q - r = 2p + q, \\ V(3) - V(2) &= p\lfloor 3\alpha \rfloor + 3q + r - p\lfloor 2\alpha \rfloor - 2q - r = p + q, \\ V(4) - V(3) &= p\lfloor 4\alpha \rfloor + 4q + r - p\lfloor 3\alpha \rfloor - 3q - r = 2p + q, \end{aligned}$$

and the result follows, since $p \neq 0$. When $\frac{4}{3} < \alpha < \frac{3}{2}$, we have $\lfloor \alpha \rfloor = 1$, $\lfloor 2\alpha \rfloor = 2$, $\lfloor 3\alpha \rfloor = 4$, and $\lfloor 4\alpha \rfloor = 5$. So this time $V(2) - V(1) = p + q$, $V(3) - V(2) = 2p + q$ and $V(4) - V(3) = p + q$, leading to the same conclusion. □

Remark 10. We note for further use that solving the equations in the proof of Lemma 9 for p and q, supplemented with an equation for r, yields in the case $\frac{3}{2} < \alpha < \frac{5}{3}$ that

$$\begin{cases} p &= -V(1) + 2V(2) - V(3) \\ q &= \ \ V(1) - 3V(2) + 2V(3) \\ r &= \ \ V(1) + V(2) - V(3). \end{cases}$$

Let $\alpha = \varphi$ be the golden mean. Then the classical solution is $(p, q, r) = (1, 0, 0)$ and $(s, t, u) = (1, 1, 0)$, which corresponds to the Beatty pair $(\lfloor n\varphi \rfloor)$, $(\lfloor n\varphi^2 \rfloor)$. Another solution is given by

$$(p, q, r) = (-1, 3, -1), \quad (s, t, u) = (1, 2, 0),$$

which corresponds to the Beatty pair $\left(\left\lfloor n\left(\frac{5-\sqrt{5}}{2}\right)\right\rfloor\right)$, $\left(\left\lfloor n\left(\frac{5+\sqrt{5}}{2}\right)\right\rfloor\right)$, that is, $(\lfloor n(3 - \varphi) \rfloor)$, $(\lfloor n(\varphi + 2) \rfloor)$.

Theorem 11. *Let $\alpha = \varphi$. Then there are exactly two nontrivial increasing solutions to the complementary pair problem*: $(p, q, r, s, t, u) = (1, 0, 0, 1, 1, 0)$ *and* $(p, q, r, s, t, u) = (-1, 3, -1, 1, 2, 0)$.

Proof. Recall that $V(1) = 1$. Note that $V(2) < 5$, since otherwise $(W(1), W(2), W(3)) = (2, 3, 4)$, which is not allowed by Lemma 9. There are therefore three cases to consider, according to the value of $V(2)$.

(1) The case $V(1) = 1$, $V(2) = 2$. Then by Lemma 9, $V(3) = 3$ is not possible.

   (a) If $V(3) = 4$, then, by Remark 10, $p = -1$, $q = 3$, $r = -1$, which is one of the two solutions.

   (b) If $V(3) = 5$, then, by Remark 10, $p = -2$, $q = 5$, $r = -2$, which implies that $V(4) = 6$, $V(5) = 7$, $V(6) = 10$. So $W(1) = 3$, $W(2) = 4$, $W(3) = 8$, which gives $s = -3$, $t = 7$, $u = -1$ (Remark 10 applied to W), implying $W(5) = 10$, which contradicts complementarity.

   (c) If $V(3) = m$ with $m > 5$, then $W(1) = 3$, $W(2) = 4$, $W(3) = 5$, which contradicts Lemma 9.

(2) The case $V(1) = 1$, $V(2) = 3$.

   (a) If $V(3) = 4$, then, by Remark 10, $p = 1$, $q = 0$, $r = 0$, which is one of the two solutions.

   (b) If $V(3) = 5$, then we obtain a contradiction with Lemma 9.

   (c) If $V(3) = 6$, then, by Remark 10, $p = -1$, $q = 4$, $r = -2$, which implies $V(5) = 10$. But we must then have $W(1) = 2$, $W(2) = 4$, $W(3) = 5$, so (Remark 10 applied to W), $s = 1$, $t = 0$, $u = 1$, which implies $W(6) = 10$, a contradiction with complementarity.

   (d) If $V(3) = m$ with $m > 6$, then we obtain a contradiction with Lemma 9, since then $W(2) = 4$, $W(3) = 5$, $W(4) = 6$.

(3) The case $V(1) = 1$, $V(2) = 4$.

   (a) If $V(3) = 5$, then, by Remark 10, $p = 2$, $q = -1$, $r = 0$, thus $V(4) = 8$; hence $W(1) = 2$, $W(2) = 3$, $W(3) = 6$. Hence, by Remark 10 applied to W, $s = -2$, $t = 5$, $u = -1$, so that $W(5) = 8 = V(4)$, which contradicts complementarity.

   (b) If $V(3) = 6$, then $W(1) = 2$, $W(2) = 3$, $W(3) = 5$. Thus, by Remark 10 applied to W, $s = -1$, $t = 3$, $u = 0$. Hence $W(4) = 6 = V(3)$, which contradicts complementarity.

   (c) If $V(3) = 7$, then we obtain a contradiction with Lemma 9.

   (d) If $V(3) = m$ with $m > 7$, it follows that $V(3) = 8$, since we have $W(1) = 2$, $W(2) = 3$, $W(3) = 5$, yielding, by Remark 10 applied to W, $W = (-A(n) + 3n) = 2, 3, 5, 6, 7, 9, 10, 12, 13, 14, \ldots$ With $V(3) = 8$, one obtains (by Remark 10) that $V(n) = -A(n) + 5n - 3$, but then $V(5) = 14 = W(10)$, i.e., V and W are not complementary. □

For $\alpha = \sqrt{2}$ the classical solution to the complementary pair problem is $V = A$, $W = A + 2\,\mathrm{Id}$, i.e., the Beatty pair given by $V(n) = \lfloor n\sqrt{2}\rfloor$, and $W(n) = \lfloor n(2+\sqrt{2})\rfloor$. As $\frac{4}{3} < \sqrt{2} < \frac{3}{2}$, we can use Lemma 9 and adapt Remark 10 to prove the following result, in the same way as Theorem 11.

Theorem 12. *Let $\alpha = \sqrt{2}$. Then there is a unique nontrivial increasing solution to the complementary pair problem*: $(p, q, r, s, t, u) = (1, 0, 0, 1, 2, 0)$.

We end this section with an example where $\{\alpha\} \notin \left(\frac{1}{3}, \frac{2}{3}\right)$.

Theorem 13. *Let $\alpha = \sqrt{8}$. Then there is a unique nontrivial increasing solution to the complementary pair problem*: $(p, q, r, s, t, u) = (1, 4, 0, -1, 4, 0)$.

Proof. Since $(4+\sqrt{8}, 4-\sqrt{8})$ is a Beatty pair, $(p, q, r, s, t, u) = (1, 4, 0, -1, 4, 0)$ is a solution to the complementary pair problem. To prove that it is unique is more involved. We fix $V(1) = 1$.

Let $\breve{\alpha} = \alpha - 2 = \sqrt{8} - 2$. Then $\breve{\alpha} \in (0, 1)$, and $\breve{\alpha}$ has the periodic continued fraction expansion $[0; \overline{1, 4}]$. It follows then from [Crisp et al. 1993], or from Corollary 9.1.6 in [Allouche and Shallit 2003] that the morphism $\sigma_{\breve{\alpha}}$ fixing the homogeneous Sturmian sequence $c_{\breve{\alpha}}$ is given by

$$\sigma_{\breve{\alpha}} : 0 \mapsto 11110, \qquad 1 \mapsto 111101.$$

Note that

$$V(n) = p\lfloor n\sqrt{8}\rfloor + qn + r = p\lfloor n(\sqrt{8}-2)\rfloor + (2p+q)n + r = p\lfloor n\breve{\alpha}\rfloor + (2p+q)n + r.$$

The difference sequence ΔV of V is therefore the fixed point of $\sigma_{\breve{\alpha}}$ on the alphabet $\{2p+q, 3p+q\}$. Since we require V to be increasing, both $2p+q$ and $3p+q$ have to be larger than 0. We split the possibilities according to the value of $3p+q$. The arguments below are based on the fact, following from the form of $\sigma_{\breve{\alpha}}$, that V starts with an arithmetic sequence of length 5, followed by an arithmetic sequence of length 6, both with common differences $3p+q$, and separated by a distance $2p+q$.

(1) The case $3p+q \geqslant 3$. If $3p+q \geq 3$, then $W(1) = 2$, $W(2) = 3$, so $W = sA + t\mathrm{Id} + u$ has to start with an arithmetic sequence of length 5 with common difference 1, i.e., $W(1), W(2), \ldots, W(5) = 2, 3, \ldots, 6$. Moreover, since $W(6) = 7$ is not possible (as it would imply $p = 0$), we have $V(2) = 7$,

which implies $V(3) = 13$. But then the second arithmetic sequence of W, which has length 6, does not fit in between $V(2)$ and $V(3)$.

(2) The case $3p+q=2$. In this case $V(1), \dots, V(5) = 1, 3, 5, 7, 9$, so $W(1), \dots, W(5) = 2, 4, 6, 8, 10$. Then either $V(6) = 11$, or $W(6) = 11$.

In the former case we must have $2p+q = V(6) - V(5) = 2$, which implies $p = 0$, which is trivial.

In the latter case $W(6), \dots, W(11) = 11, 13, 15, 17, 19, 21$, and $W(12) = 22$, since $W(12) - W(11) = W(6) - W(5) = 1$. But also, $V(6), \dots, V(11)$ equals $12, 14, \dots, 22$. So V and W are not complementary.

(3) The case $3p+q=1$. In this case $V(1), \dots, V(5) = 1, 2, 3, 4, 5$, so $W(1) = 6$, since $V(6) = 6$ would imply $p = 0$. Then either $V(6) = 7$, or $W(2) = 7$.

In the former case, $V(6), \dots, V(11) = 7, 8, 9, 10, 11, 12$, and $W(2) = 13$. This implies that $2 = V(6) - V(5) = 2p+q$, which leads to $(p, q, r) = (-1, 4, 0)$, and $(s, t, u) = (1, 4, 0)$, which is the announced solution.

In the latter case $W(1), \dots, W(5) = 6, 7, 8, 9, 10$, and $V(6) = 11$. So $2p+q = V(6) - V(5) = 5$. This implies $(p, q, r) = (-5, 16, -5)$, and $(s, t, u) = (-6, 19, -1)$. But then $V(12) = 22$, and $W(11) = 22$. So V and W are not complementary. □

2.1. *Generalized Pell equations.* If V and W are not increasing, then an analysis as in the proof of Theorem 11 is still possible, but very lengthy. We therefore consider another approach in this subsection. Considering the densities of V and W in $\mathbb{N}$, one sees that a *necessary* condition for $(pA + q\,\mathrm{Id} + r)$ and $(sA + t\,\mathrm{Id} + u)$ to be a complementary pair is that

$$\frac{1}{p\alpha + q} + \frac{1}{s\alpha + t} = 1 \tag{2}$$

In what follows we concentrate on the case $\alpha = \varphi = (1+\sqrt{5})/2$, but our arguments can be generalized to the case of arbitrary quadratic irrationals.

Proposition 14. *A necessary condition for the pair $V = pA + q\,\mathrm{Id} + r$ and $W = sA + t\,\mathrm{Id} + u$ to be a complementary pair is that $p \neq 0$ is a solution to the generalized Pell equation*

$$5p^2x^2 - 4x = y^2, \quad x, y \in \mathbb{Z}.$$

Proof. Using $\varphi^2 = 1 + \varphi$, a straightforward manipulation shows that (2) implies

$$(ps + pt + qs - p - s)\varphi = q + t - ps - qt.$$

But since φ is irrational, this can only hold if

$$ps + pt + qs - p - s = 0, \quad q + t - ps - qt = 0. \tag{3}$$

The first equation gives $pt = p - (p+q-1)s$. Eliminating pt from $p^2s + (q-1)pt - pq = 0$, we obtain $p^2s + (p - (p+q-1)s)(q-1) - pq = 0$. This gives the quadratic equation

$$s\,q^2 + (p-2)s\,q - (p^2 + p - 1)s + p = 0.$$

Since q is an integer, $\Delta := (p-2)^2s^2 + 4s((p^2+p-1)s - p)$ has to be an integer squared. Trivial manipulations yield that

$$\Delta = 5p^2s^2 - 4ps. \tag{4}$$

Since p divides the square Δ, $5p^2s^2 - 4ps = p^2y^2$ for some integer y, and hence p also divides s. If we put $s = px$, we obtain $5p^3x^2 - 4p^2x = p^2y^2$, which finishes the proof of the proposition. □

Actually there is a simple characterization of the integers p such that the diophantine equation above has a solution.

Proposition 15. *The generalized Pell equation*

$$5p^2x^2 - 4x = y^2, \quad x, y \in \mathbb{Z},$$

has a solution for $p > 0$ if and only if p divides some Fibonacci number of odd index, i.e., if and only p divides some number in the set $\{1, 2, 5, 13, 34, \ldots\}$.

Proof. First suppose that there are integers $p > 0$ and $x, y \in \mathbb{Z}$ such that $5p^2x^2 - 4x = y^2$. Let $d := \gcd(x, y)$ and $x' = x/d$, $y' = y/d$, so that $\gcd(x', y') = 1$. We thus have

$$5p^2dx'^2 - 4x' = dy'^2.$$

Thus x' divides dy'^2, but it is prime to y', hence x' divides d. Since clearly d divides $4x'$, we have $d = \alpha x'$ for some α dividing 4, hence α belongs to $\{1, 2, 4\}$. This yields $\alpha(5p^2x'^2 - y'^2) = 4$. We distinguish three cases.

(1) If $\alpha = 1$, then we have $5p^2x'^2 - y'^2 = 4$. But the equation $5X^2 - 4 = Y^2$ has an integer solution if and only if X is a Fibonacci number with odd index [Lind 1968, p. 91]. Hence px' must be a Fibonacci number with odd index, thus p divides a Fibonacci number with odd index.

(2) If $\alpha = 2$, then we have $5p^2x'^2 - y'^2 = 2$. Note that x' must be odd, otherwise x' and y' would be even, which contradicts $\gcd(x', y') = 1$. Thus $5p^2x'^2 \equiv p^2 \bmod 4$, hence $p^2 - 2 \equiv y'^2 \bmod 4$. If p is even, this yields $y'^2 \equiv 2 \bmod 4$, while if p is odd, this gives $y'^2 \equiv 3 \bmod 4$. There is no such y' in both cases.

(3) If $\alpha = 4$, then we have $5p^2x'^2 - y'^2 = 1$, thus $5(2px')^2 - (2y')^2 = 4$, then $2px'$ must be a Fibonacci number with odd index, thus p divides a Fibonacci number with odd index.

Now suppose that p divides some Fibonacci number with odd index, say there exists a k with $F_{2k+1} = p\beta$. We will construct an integer solution in (x, y) to the equation $5p^2x^2 - 4x = y^2$. We know (again [Lind 1968, p. 91]) that there exists some integer γ with $5F_{2k+1}^2 - 4 = \gamma^2$ thus $5p^2\beta^2 - 4 = \gamma^2$. Let $x = \beta^2$ and $y = \beta\gamma$. Then

$$5p^2x^2 - 4x = 5p^2\beta^4 - 4\beta^2 = \beta^2(5p^2\beta^2 - 4) = \beta^2\gamma^2 = y^2. \qquad \square$$

Corollary 16. *If -1 is not a square modulo p, there are no solutions to the complementary pair problem. This is in particular the case if p has a prime divisor congruent to* 3 *modulo* 4.

(The sequence of such integers p, which starts with $1, 2, 5, 10, 13, 17, 25, 26, 29, 34, 37, 41, \ldots$, is labeled A008784 in *The on-line encyclopedia of integer sequences* [OEIS].)

Proof. We will prove that if there are solutions to the complementary problem for p, thus if p divides an odd-indexed Fibonacci number (Propositions 14 and 15), then -1 is a square modulo p. Using again the characterization in [Lind 1968, p. 91], there exist two integers x, y with $5p^2x^2-4=y^2$. We distinguish two cases.

(i) If p is odd, we have $y^2 \equiv -4 \bmod p$ and $2^2 \equiv 4 \bmod p$. But 2 is invertible modulo p, hence, by taking the quotient of the two relations, we obtain that -1 is a square modulo p.

(ii) If p is even, remembering that $px = F_{2k+1}$ for some k, we claim that p must be congruent to 2 modulo 4 and that x must be odd. Namely the sequence of odd-indexed Fibonacci numbers, reduced modulo 4, is easily seen to be the periodic sequence $(1\ 2\ 1)^\infty$. Hence it never takes the value 0 modulo 4. The equality $5p^2x^2-4=y^2$ implies that y must be even, thus we have $5(p/2)^2x^2-1=(y/2)^2$, say $(y/2)^2=-1+z(p/2)$. Up to replacing $(y/2)$ with $(y+p)/2$, we may suppose that $(y/2)$ is even (recall that $p/2$ is odd). Thus $z(p/2)$ is even, hence z is even, say $z=2z'$. This gives $(y/2)^2=-1+z'p$; thus -1 is a square modulo p. □

Remark 17. We have just seen that if the integer p divides some odd-indexed Fibonacci number (OEIS A008784) then -1 is a square modulo p. A natural question is then whether it is true that if -1 is a square modulo p, then p must divide some odd-indexed Fibonacci number. The answer is negative, since on one hand $12^2 \equiv -1 \bmod 29$, and, on the other hand, the sequence of odd-indexed Fibonacci numbers modulo 29 is the periodic sequence $(1, 2, 5, 13, 5, 2, 1)^\infty$ which is never zero.

Let us look at examples of solutions to the diophantine equation for values of p that divide some Fibonacci number with odd index. Consider, for example, the case where $p=s$. Then equation (4) becomes $\Delta = 5p^4-4p^2$, so the diophantine equation is

$$5x^2-4=y^2, \quad x, y \in \mathbb{Z}.$$

For $p=F_1=1$ we obtain the two sequences $V=A+r$ and $W=A+\mathrm{Id}+u$. These are complementary only when $r=u=0$, and we obtain the classical Beatty pair $(A, A+\mathrm{Id})$.

For $p=F_3=2$ we obtain the two sequences $V=2A+2\,\mathrm{Id}+r$ and $W=2A-2\,\mathrm{Id}+u$. These cannot be complementary for any r and u, since for $u=0$ we have $W(n)=2\lfloor n\varphi\rfloor-2n=2\lfloor n(\varphi-1)\rfloor$, which gives all even numbers, since $\varphi-1<1$. This an example where equation (2) does not apply, since W as a function is not injective.

For $p=F_5=5$ we obtain the two sequences $V=5A+4\,\mathrm{Id}+r$ and $W=5A-7\,\mathrm{Id}+u$. To make these complementary we are forced to choose $r=u=3$, and we obtain

$$V=(12, 26, 35, 49, 63, 72, 86, 95, 109, 123, 132, 146, 160, 169, 183, 192, 206, 220, 229, 243, 252, 266, \dots),$$

$$W=(1, 4, 2, 5, 8, 6, 9, 7, 10, 13, 11, 14, 17, 15, 18, 16, 19, 22, 20, 23, 21, 24, 27, 25, 28, 31, 29, 32, 30, \dots).$$

Now a proof that V and W form a complementary pair is much harder when we let V start with $V(0)=3$, to include 3 in the union. We can perform the following trick. We split W into $(W(A(n)))_{n\geq1}$, and $(W(B(n)))_{n\geq1}$ (cf. Corollary 2). The two sequences WA and WB are increasing, and we can prove that $(V(n))_{n\geq0}$, $(W(A(n)))_{n\geq1}$, and $(W(B(n)))_{n\geq1}$ form a partition of the positive integers by proving that the three-letter sequence obtained by applying the morphism $0 \mapsto 1120$, $1 \mapsto 11100$ to the fixed point of the morphism g given by $g: 0 \mapsto 01,\ 1 \mapsto 011$, has the property that the preimages of 0, 1 and 2 are precisely these three sequences. See Theorem 25 and its proof for a similar result.

For $p = F_{2m+1} \geq 13$ it seems that we can always choose r and u for in such a way that we get almost complementary sequences: namely, e.g., for $p = 13$ we find $q = 9$ and $t = -20$. If we take $r = u = 9$, then we almost get a complementary pair. One finds $V = 9, 31, 66, 88, 123, 158, 180, 215, \dots$ and $W = 2, 8, 1, 7, 13, 6, 12, 5, 11, 17, 10 \dots$. So 3 and 4 are missing. We thought we could prove, perhaps using something like the Lambek–Moser Theorem [Lambek and Moser 1954], that for all $F_{2m+1} > 5$ the two sequences are complementary excluding finitely many values, but we were not successful.

3. Complementary triples

Here we will find several complementary triples consisting of sequences $V_i = p_i A + q_i \operatorname{Id} + r_i$, $i = 1, 2, 3$, where $A(n) = \lfloor n\alpha \rfloor$, and α is a real number.

It is interesting that the case $p_1 = p_2 = p_3 = 1$ cannot be realized. This was proved by Uspensky in 1927; see [Fraenkel 1977].

The case with different α_i was analyzed in [Tijdeman 1996] for *rational* α_i, $1 = 1, 2, 3$. Also see [Tijdeman 2000] for the inhomogeneous Beatty case $(V_i(n))_n = (\lfloor n\alpha_i + \beta_i \rfloor)_n$, $i = 1, 2, 3$.

There is one triple in which we will be particularly interested (see Theorem 25):

$$(p_1, q_1, r_1) = (2, -1, 0), \quad (p_2, q_2, r_2) = (4, 3, 2), \quad (p_3, q_3, r_3) = (2, -1, 2).$$

We allow the sequences (V_i) to be each indexed either by $\{0, 1, 2, \dots\}$ or by $\{1, 2, \dots\}$.

3.1. ***Two classical triples.*** In this subsection α is always the golden mean φ. Let once more $A(n) = \lfloor n\varphi \rfloor$ for $n \geq 1$ be the terms of the lower Wythoff sequence, and let B be given by $B(n) = \lfloor n\varphi^2 \rfloor$ for $n \geq 1$, the upper Wythoff sequence. Then we have the disjoint union

$$A(\mathbb{N}) \cup B(\mathbb{N}) = \mathbb{N}. \tag{5}$$

Since $B = A + \operatorname{Id}$, this is the classical complementary pair $((1, 0, 0), (1, 1, 0))$.

Here is a way to create complementary triples from complementary pairs.

Proposition 18. *Let (V, W) be a golden mean complementary pair $V = pA + q \operatorname{Id} + r$ and $W = sA + t \operatorname{Id} + u$. Then (V_1, V_2, V_3) is a complementary triple, where the three parameters of V_1 are $(p + q, p, r - p)$, those of V_2 are $(2p + q, p + q, r)$, and $V_3 = W$.*

Proof. Substituting (5) in $V(\mathbb{N}) \cup W(\mathbb{N}) = \mathbb{N}$ we obtain the disjoint union

$$V(A(\mathbb{N})) \cup V(B(\mathbb{N})) \cup W(\mathbb{N}) = \mathbb{N}.$$

Then Corollary 2 implies the statement of the proposition. □

Remark 19. Actually Proposition 18 and Corollary 2 can be generalized to cover an infinite family of quadratic irrationals, but their statements will not be true for all quadratic irrationals. We hope to revisit this point in a future article.

Applying Proposition 18 to the basic complementary pair $((1, 0, 0), (1, 1, 0))$ gives that

$$((1, 1, -1), (2, 1, 0), (1, 1, 0)) \quad \text{and} \quad ((1, 0, 0), (2, 1, -1), (3, 2, 0))$$

are complementary triples, which we will call classical triples. (The corresponding sequences are indexed in OEIS as (A003623, A003622, A001950) and (A000201, A035336, A101864). The first classical triple is given at the end of [Skolem 1957].)

Let $w = 1231212312312\ldots$ be the fixed point of the morphism

$$1 \mapsto 12,\ 2 \mapsto 3,\ 3 \mapsto 12.$$

Then $w^{-1}(1) = AA,\ w^{-1}(2) = B$ and $w^{-1}(3) = AB$ give the three sequences V_1, V_3 and V_2 of the first classical triple (see [Dekking 2016]).

The question arises: is there also a morphism generating the second triple? The answer is positive.

Proposition 20. *Let* $(V_1, V_2, V_3) = (A,\ 2A + \mathrm{Id}{-}1,\ 3A + 2\mathrm{Id}) = (A,\ BA,\ BB)$. *Then* (V_1, V_2, V_3) *is a complementary triple. Let* μ *be the morphism on* $\{1, 2, 3\}$ *given by*

$$\mu : 1 \mapsto 121,\ 2 \mapsto 13,\ 3 \mapsto 13,$$

with fixed point z. *Then* $z^{-1}(1) = V_1,\ z^{-1}(2) = V_2$ *and* $z^{-1}(3) = V_3$.

Proof. The four words of length 3 occurring in the infinite Fibonacci word x_{F} are 010, 100, 001, 101. Coding these with the alphabet $\{1, 2, 3, 4\}$ in the given order, they generate the 3-block morphism $\hat{f}_3$ that describes the successive occurrences of the words of length 3 in x_{F} (cf. [Dekking 2016]). It is given by

$$\hat{f}_3(1) = 12, \quad \hat{f}_3(2) = 3, \quad \hat{f}_3(3) = 14, \quad \hat{f}_3(4) = 3.$$

It has just one fixed point, which is

$$z' := 1, 2, 3, 1, 4, 1, 2, 3, 1, 2, 3, 1, 4, 1, 2, 3, \ldots.$$

We claim that

$$z'^{-1}(1) = AA,\ z'^{-1}(2) = BA, \quad z'^{-1}(3) = AB, \quad z'^{-1}(4) = BB.$$

To see this, note that the 3-block 010 in x_{F} uniquely decomposes as $010 = f(0)0$. It follows that the m^{th} occurrence of 010 in x_{F} corresponds exactly to the m^{th} occurrence of 0 in $f^{-1}(x_{\mathrm{F}}) = x_{\mathrm{F}}$. This implies that the positions of the occurrences of 010 are of the form $A(A(n))$, and also that the occurrences of 001 are of the form $A(B(n))$, since $B(\mathbb{N})$ is the complement of $A(\mathbb{N})$.

For the 3-block 100, we note that it always occurs in x_{F} as factor of 0100, which uniquely decomposes in x_{F} as $0100 = f^2(0)0$. It follows that the m^{th} occurrence of 100 in x_{F} corresponds exactly to the m^{th} occurrence of 0 in $f^{-2}(x_{\mathrm{F}}) = x_{\mathrm{F}}$. This implies that the positions of the occurrences of 100 are of the form $B(A(n))$, and also that the occurrences of 101 are of the form $B(B(n))$.

Since the 0's in x_{F} occur either as prefix of 001, or of 010, we see that we have to merge the letters 1 and 3 to obtain the sequence A. This is not possible with $\hat{f}_3$. However, the square of this 3-block morphism is given by

$$1 \mapsto 123,\ 2 \mapsto 14,\ 3 \mapsto 123,\ 4 \mapsto 14,$$

and now we *can* consistently merge 1 and 3 to the single letter 1, obtaining the morphism μ, after mapping 4 to 3. Under this projection the sequence z' maps to z. □

3.2. ***Nonclassical triples.*** Let $\mathcal{L}$ be a language, i.e., a sub-semigroup of the free semigroup generated by a finite alphabet under the concatenation operation. A homomorphism of $\mathcal{L}$ into the natural numbers is a map $\mathrm{S} : \mathcal{L} \to \mathbb{N}$ satisfying $\mathrm{S}(vw) = \mathrm{S}(v) + \mathrm{S}(w)$, for all $v, w \in \mathcal{L}$.

Let $\mathcal{L}_\mathrm{F}$ be the Fibonacci language, i.e., the set of all words occurring in the Fibonacci word x_F, the iterative fixed point of the morphism f defined on $\{0, 1\}^*$ by $f : 0 \mapsto 01,\ 1 \mapsto 0$.

Theorem 21 [Dekking 2018a]. *Let* $\mathrm{S} : \mathcal{L}_\mathrm{F} \to \mathbb{N}$ *be a homomorphism. Define* $a = \mathrm{S}(0)$, $b = \mathrm{S}(1)$. *Then* $\mathrm{S}(\mathcal{L}_\mathrm{F})$ *is the union of the two generalized Beatty sequences*

$$\big((a-b)\lfloor n\varphi\rfloor + (2b-a)n\big) \quad \textit{and} \quad \big((a-b)\lfloor n\varphi\rfloor + (2b-a)n + a - b\big).$$

For a few choices of a and b, the two sequences in $\mathrm{S}(\mathcal{L}_\mathrm{F})$ and the sequence $\mathbb{N} \setminus \mathrm{S}(\mathcal{L})$ form a complementary triple of generalized Beatty sequences. The goal of this section is to prove this for $a = 3,\ b = 1$. It turns out that the three sequences

$$(2\lfloor n\varphi\rfloor - n)_{n\ge1},\ (2\lfloor n\varphi\rfloor - n + 2)_{n\ge1},\ (4\lfloor n\varphi\rfloor + 3n + 2)_{n\ge0},$$

form a complementary triple.

Remark 22. Note that the indices for $(4\lfloor n\varphi\rfloor) + 3n + 2)_{n\ge0}$ are $(n \ge 0)$, not $(n \ge 1)$.

It is easy to see that the Fibonacci word x_F can be obtained as an infinite concatenation of two kinds of blocks, namely 01 and 001 (part (i) of Lemma 23 below). Kimberling introduced in the OEIS the sequence A284749 obtained by replacing every block 001 in this concatenation by 2. We let $x_\mathrm{K} = A284749$ denote this sequence.

Lemma 23. *Let* f, g, h, k *be the morphisms defined on* $\{0, 1\}^*$ *by*

$$f : 0 \mapsto 01,\ 1 \mapsto 0; \quad g : 0 \mapsto 01,\ 1 \mapsto 011; \quad h : 0 \mapsto 01,\ 1 \mapsto 001; \quad k : 0 \mapsto 01,\ 1 \mapsto 2.$$

The following equalities hold:

(i) $x_\mathrm{F} = f^\infty(0) = h(g^\infty(0))$.

(ii) $x_\mathrm{K} = k(g^\infty(0))$.

Proof. (i) An easy induction proves that for all $n \ge 0$ one has $hg^k = f^{2k}h$. (Note that it suffices to prove that the values of both sides are equal when applied to 0 and to 1.) By letting n tend to infinity this implies $hg^\infty(0) = f^\infty(0)$.

(ii) Assertion (i) gives that x_F is an infinite concatenation of blocks $h(0) = 01$ and $h(1) = 001$, obtained as image under h from $g^\infty(0)$. So substituting 2 for 001 in x_F is the same as substituting 01 for 0 and 2 for 1 in $g^\infty(0)$. □

It is interesting that x_K is fixed point of a morphism i, given by $i : 0 \mapsto 01,\ 1 \mapsto 2,\ 2 \mapsto 0122$. This follows from the relation $kg^n = i^{n+1}$ for all n, which is easily proved by induction.

Lemma 24. *Define the morphism* ℓ *from* $\{0, 1\}^*$ *to* $\{0, 1, 2\}^*$ *by* $\ell : 0 \mapsto 012,\ 1 \mapsto 0022$. *Then the sequence* $v = (v_n)_{n\ge1} = \ell(g^\infty(0))$ *is obtained from* x_K *by replacing* 1 *by* 0 *in all blocks* 0122 (*but not in* 0120).

Proof. Note that $kg : 0 \mapsto 012,\ 1 \mapsto 0122$. Lemma 24 then follows from $x_\mathrm{K} = k(g^\infty(0)) = kg(g^\infty(0))$. □

Theorem 25. *Let v be the sequence defined above, i.e., $v = \ell(g^\infty(0))$, where g and ℓ are the morphisms defined by $g : 0 \mapsto 01$, $1 \mapsto 011$ and $\ell : 0 \mapsto 012$, $1 \mapsto 0022$. Then the increasing sequences of integers defined by $v^{-1}(0)$, $v^{-1}(1)$, $v^{-1}(2)$ form a partition of the set of positive integers $\mathbb{N}^*$. Furthermore*:

- $v^{-1}(0) = \{1, 4, 5, 8, 11, 12, 15, 16, 19, 22, \ldots\}$ *is equal to the sequence of integers* $(2\lfloor n\varphi\rfloor - n)_{n\geq 1}$, *where φ is the golden ratio* $\frac{1+\sqrt{5}}{2}$ (OEIS A050140).
- $v^{-1}(1) = \{2, 9, 20, 27, \ldots\}$ *is equal to the sequence of integers* $(4\lfloor n\varphi\rfloor + 3n + 2)_{n\geq 0}$.
- $v^{-1}(2) = \{3, 6, 7, 10, 13, 14, 17, 18, 21, 24, \ldots\}$ *is equal to the sequence* $(2\lfloor n\varphi\rfloor - n + 2)_{n\geq 1}$ (that is, $2+$ A050140).

Proof. We see from Lemma 24 that the positions of 2 in v are the same as the positions of 2 in x_{K}. In Section 4 it is proved that $x_{\mathrm{K}}^{-1}(2) = (2\lfloor n\varphi\rfloor - n + 2)$, see Example 30, and so the third assertion of the theorem follows.

Inspection of the occurrences of 0 and 2 in $\ell(0)$ and $\ell(1)$ then shows the first assertion to be true too.

For the proof of the second assertion consider $\ell(01) = 0\mathbf{1}20022$, $\ell(011) = 0\mathbf{1}200220022$. Since $g^\infty(0)$ is a concatenation of the words $g(0) = 01$ and $g(1) = 011$, we see from this that the differences of indices of the positions where 1's in v occur are 7 or 11, and moreover, that a 0 in $g^\infty(0)$ generates a difference 7, a 1 in $g^\infty(0)$ generates a difference 11.

It is well known and easy to prove that $g^\infty(0)$ equals the binary complement of the Fibonacci word prefixed with the letter 1. From this it follows that $\Delta v^{-1}(1)$ is the Fibonacci word on the alphabet $\{11, 7\}$. Now Lemma 8 gives the generalized Beatty sequence $V = (4\lfloor n\varphi\rfloor + 3n + 2)$. The first element 2 in $v^{-1}(1)$ is obtained by letting V start at $n = 0$ instead of $n = 1$. □

Remark 26. Some of the sequences above are images of Sturmian sequences by a morphism. Namely $v = \ell(g^\infty(0))$, $x_{\mathrm{K}} = k(g^\infty(0))$. Such sequences are examples of sequences called *quasi-Sturmian* in [Cassaigne 1998]. Their block complexity is of the form $n + C$ for n large enough ($C = 1$ for Sturmian sequences). These sequences were studied, e.g., in [Paul 1974/75; Coven 1974/75; Cassaigne 1998].

4. Generalized Beatty sequences and return words

In this section we show that generalized Beatty sequences are closely related to return words.

Theorem 27. *Let x_{F} be the Fibonacci word, and let w be any word in the Fibonacci language $\mathcal{L}_{\mathrm{F}}$. Let Y be the sequence of positions of the occurrences of w in x_{F}. Then Y is a generalized Beatty sequence, i.e., for all $n \geq 0$, $Y(n+1) = p\lfloor n\varphi\rfloor + qn + r$ with parameters p, q, r, which can be explicitly computed.*

Proof. Let $x_{\mathrm{F}} = r_0(w)r_1(w)r_2(w)r_3(w)\ldots$, written as a concatenation of return words of the word w (cf. [Huang and Wen 2015], Lemma 1.2). According to Theorem 2.11 in [Huang and Wen 2015], if we skip $r_0(w)$, then the return words occur as the Fibonacci word on the alphabet $\{r_1(w), r_2(w)\}$. Thus the distances between occurrences of w in x_{F} are equal to $l_1 := |r_1(w)|$ and $l_2 := |r_2(w)|$. We can apply Lemma 8, which yields $p = l_1 - l_2$ and $q = 2l_2 - l_1$. Inserting $n = 0$, we find that $r = |r_0(w)| + 1$, as the first occurrence of w is at the beginning of $r_1(w)$. □

4.1. ***The Kimberling transform.*** Here we will obtain nonclassical triples appearing in another way, namely as the three indicator functions $x^{-1}(0)$, $x^{-1}(1)$ and $x^{-1}(2)$, of a sequence x on an alphabet

$\{0, 1, 2\}$ of three symbols. In our examples the sequence x is a transform $\mathcal{T}(x_{\mathrm{F}})$ of the Fibonacci word $x_{\mathrm{F}} = 01001010010010100\ldots$. These transforms have been introduced by Kimberling. Our main example is $\mathcal{T} : [001 \mapsto 2]$. Replacing each 001 in $x_{\mathrm{F}} = 01001010010010100\ldots$ by 2 gives $x_{\mathrm{K}} = 01201220120\ldots$.

For the transform method $\mathcal{T}$ we can derive a general result similar to Theorem 27. However, since Kimberling applies the StringReplace procedure from Mathematica, which replaces occurrences of w consecutively from left to right, we do not obtain a sequence of return words in the case that w has overlaps in x_{F}. This restricts the number of words w to which Theorem 29 below applies.

Definition 28 [Wen and Wen 1994, pp. 593–594]. Let w be a factor of the Fibonacci word x_{F}. We say that w has an *overlap* in x_{F} if there exist nonempty words x, y and z such that $w = xy = yz$, and the word xyz is a factor of the Fibonacci word.

Theorem 29. *Let x_{F} be the Fibonacci word, and let w be a factor of x_{F} that has no overlap in x_{F}. Consider the transform $\mathcal{T}(x_{\mathrm{F}})$, which replaces every occurrence of the word w in x_{F} by the letter 2. Let Y be the sequence $(\mathcal{T}(x_{\mathrm{F}}))^{-1}(2)$, i.e., the positions of 2's in $\mathcal{T}(x_{\mathrm{F}})$. Then Y is a generalized Beatty sequence, i.e., for all $n \geq 1$, $Y(n) = p\lfloor n\varphi\rfloor + qn + r$, with parameters p, q, r, which can be explicitly computed.*

Proof. As in the proof of Theorem 27, let $x_{\mathrm{F}} = r_0(w)r_1(w)r_2(w)\ldots$, written as a concatenation of return words of the word w. Now the distances between 2's in $\mathcal{T}(x_{\mathrm{F}})$ are equal to $l_1 := |r_1(w)| - |w| + 1$ and $l_2 := |r_2(w)| - |w| + 1$. We can apply Lemma 8, which gives $p = l_1 - l_2$, $q = 2l_2 - l_1$. Inserting $n = 1$, we find that $r = |r_0(w)| - l_2 + 1$. □

Example 30. We take $\mathcal{T} : [001 \mapsto 2]$, with image $\mathcal{T}(x_{\mathrm{F}}) = 01201220120\ldots$, so $Y = (3, 6, 7, 10, \ldots)$. Here $r_0(w) = 01$, $r_1(w) = 00101$, $r_2(w) = 001$. This gives $l_1 = 3$, $l_2 = 1$, implying $p = 2$, $q = -1$ and $r = 2$. So Y is the generalized Beatty sequence $(Y_n)_{n\geq 1} = (2\lfloor n\varphi\rfloor - n + 2)_{n\geq 1}$.

The question arises whether not only $\mathcal{T}(x_{\mathrm{F}})^{-1}(2)$, but also $\mathcal{T}(x_{\mathrm{F}})^{-1}(0)$ and $\mathcal{T}(x_{\mathrm{F}})^{-1}(1)$ are generalized Beatty sequences. In general this is not true. However, this holds for $\mathcal{T} : [001 \mapsto 2]$.

Theorem 31. *Let $\mathcal{T} : [001 \mapsto 2]$, and let $x_{\mathrm{K}} := \mathcal{T}(x_{\mathrm{F}})$. Then the three sequences $x_{\mathrm{K}}^{-1}(0)$, $x_{\mathrm{K}}^{-1}(1)$, $x_{\mathrm{K}}^{-1}(2)$ form a complementary triple of generalized Beatty sequences.*

Proof. According to Example 30 we have that $x_{\mathrm{K}}^{-1}(2) = (2\lfloor n\varphi\rfloor - n + 2)_{n\geq 1}$. Since clearly $x_{\mathrm{K}}^{-1}(0) = x_{\mathrm{K}}^{-1}(1) - 1$, our remaining task is to prove that $(Z(n))_{n\geq 0} := x_{\mathrm{K}}^{-1}(1)$ is a generalized Beatty sequence. The return word structure of the word $w = 001$ in x_{F} is given by

$$r_0(w) = 01, \quad r_1(w) = 00101, \quad r_2(w) = 001.$$

Note that $Z(0) = 2$, the 1 coming from $r_0(w)$. This is exactly the reason why it is convenient to start Z from index 0: the other 1's are coming from the $r_1(w)$'s — note that $r_2(w)$ is mapped to 2.

The differences between the indices of occurrences of 2 in x_{K} are given by the Fibonacci word $3133131331\ldots$, which codes the appearance of the words $r_1(w)$ and $r_2(w)$. Therefore, to obtain the differences between the indices of occurrences of 1 in x_{K}, we have to map the word $w' = 13$ to 4, obtaining the word $u = 343443\ldots$. To obtain a description of u, we apply Theorem 27 a second time with $w' = 13$. We have $r_0(w') = 3$, $r_1(w') = 133$, $r_2(w') = 13$. So $l_1 = |r_1(w')| - |w'| + 1 = 2$, and $l_2 = |r_2(w')| - |w'| + 1 = 1$, which gives $p = l_1 - l_2 = 1$, $q = 2l_2 - l_1 = 0$. The conclusion is that positions of 4 in u are given by the generalized Beatty sequence $(\lfloor n\varphi \rfloor + 1)_{n\geq 1}$. This forces that u is nothing else than the Fibonacci word on $\{4, 3\}$, preceded by 3. But then Z is a generalized Beatty sequence with parameters $p = 1$, $q = 2$. Since $Z(1) = 5$, we must have $r = 2$, which happens to fit perfectly with the value $Z(0) = 2$. □

Here is an example where $\mathcal{T}(x_F)^{-1}(0)$ and $\mathcal{T}(x_F)^{-1}(1)$ are *not* generalized Beatty sequences.

Example 32. We take $\mathcal{T} : [00100 \mapsto 2]$, with image $\mathcal{T}(x_F) = 010010121010010121012\ldots$, so $Y = (8, 17, 21\ldots)$. Here $r_0(w) = 0100101$, $r_1(w) = 0010010100101$, $r_2(w) = 00100101$. This gives $l_1 = 9$ and $l_2 = 4$, so $p = 5$ and $q = -1$ and $r = 4$. Therefore Y is the generalized Beatty sequence $(Y_n)_{n\geq 1} = (5\lfloor n\varphi \rfloor - n + 4)_{n\geq 1}$. The positions of 0 are given by $(\mathcal{T}(x_F))^{-1}(0) = 1, 3, 4, 6, 10, 12, 13, \ldots$, with difference sequence $2, 1, 2, 4, 2, 1, \ldots$, so by Lemma 8 this sequence is not a generalized Beatty sequence. However, it can be shown that $(\mathcal{T}(x_F))^{-1}(0)$ is a union of four generalized Beatty sequences, and the same holds for $(\mathcal{T}(x_F))^{-1}(1)$.

Here is the general result.

Theorem 33. *For a nonoverlapping word w from the Fibonacci language let $\mathcal{T} : [w \mapsto 2]$, and let $Y := \mathcal{T}(x_F)$. Suppose w satisfies*

$$|r_0(w)| \leq |r_1(w)| - |w|. \tag{SR0}$$

Then the three sequences $Y^{-1}(0)$, $Y^{-1}(1)$, $Y^{-1}(2)$ are finite unions of generalized Beatty sequences.

Note that we already know by Theorem 29 that $Y^{-1}(2)$ is a single generalized Beatty sequence. Condition (SR0) states that $r_0(w)$ is short relative to $r_1(w)$.

For the proof of Theorem 33 one needs the following proposition.

Proposition 34. *Let w be a word from the Fibonacci language, and let $r_0(w)r_1(w)r_2(w)\ldots$ be the return sequence of w in the Fibonacci word x_F. Then:*

(i) *$r_0(w)$ is a suffix of $r_1(w)$.*

(ii) *If $r_2(w) = wt_2(w)$, then $t_2(w)$ is a suffix of $r_1(w)$.*

Proof. Let $s_0 = 1, s_1 = 00, s_2 = 101, s_3 = 00100, \ldots$ be the singular words introduced in [Wen and Wen 1994]. According to [Huang and Wen 2015, Theorem 1.9] there is a unique largest singular word s_k occurring in w, so we can write $w = \mu_1 s_k \mu_2$, for two words μ_1, μ_2 from the Fibonacci language. It is known — see [Wen and Wen 1994] and the remarks after [Huang and Wen 2015, Proposition 1.6] — that the two return words of the singular word s_k are

$$r_1(s_k) = s_k s_{k+1}, \quad r_2(s_k) = s_k s_{k-1}.$$

According to [Huang and Wen 2015, Lemma 3.1], the two return words of w are given by

$$r_1(w) = \mu_1 r_1(s_k)\mu_1^{-1}, \quad r_2(w) = \mu_1 r_2(s_k)\mu_1^{-1}.$$

Substituting the first equation in the second, we obtain the key equation

$$r_1(w) = \mu_1 s_k s_{k+1}\mu_1^{-1}, \quad r_2(w) = \mu_1 s_k s_{k-1}\mu_1^{-1}. \tag{6}$$

Proof of (i): We compare the return word decompositions of x_{F} by s_k and by w:

$$\begin{aligned} r_0(s_k)r_1(s_k)r_2(s_k)r_1(s_k)\dots &= r_0(w)r_1(w)r_2(w)r_1(w)\dots \\ &= r_0(w)\mu_1 r_1(s_k)\mu_1^{-1}\mu_1 r_2(s_k)\mu_1^{-1}\mu_1 r_1(s_k)\mu_1^{-1}\dots. \end{aligned}$$

It follows that $r_0(s_k) = r_0(w)\mu_1$, and so $r_0(w) = r_0(s_k)\mu_1^{-1}$. By [Huang and Wen 2015, Lemma 2.3], $r_0(s_k)$ equals s_{k+1}, with the first letter deleted. Thus we obtain from (6) that $r_0(w)$ is a suffix of $r_1(w)$.

Proof of (ii): Since $s_{k+1} = s_{k-1}s_{k-3}s_{k-1}$, by [Wen and Wen 1994, Property 2], we can make the following computation, starting from (6):

$$r_1(w) = \mu_1 s_k s_{k+1}\mu_1^{-1} = w\mu_2^{-1}s_{k+1}\mu_1^{-1} = w\mu_2^{-1}s_{k-1}s_{k-3}s_{k-1}\mu_1^{-1} = w\mu_2^{-1}s_{k-1}s_{k-3}\mu_2\mu_2^{-1}s_{k-1}\mu_1^{-1}.$$

For $r_2(w)$ we have

$$r_2(w) = \mu_1 s_k s_{k-1}\mu_1^{-1} = w\mu_2^{-1}s_{k-1}\mu_1^{-1}.$$

Now note that in this concatenation μ_2^{-1} cancels against a suffix of w. We claim that it also cancels against a prefix of s_{k-1}. This follows, since by [Huang and Wen 2015, Proposition 2.5] *any* occurrence of s_k in x_{F} is directly followed by a $s_{k+1} = s_{k-1}s_{k-3}s_{k-1}$ with the last letter deleted. It now follows that $t_2(w) = \mu_2^{-1}s_{k-1}\mu_1^{-1}$, and we see that this word is a suffix of $r_1(w)$. □

Proof of Theorem 33. In view of property (ii) in Proposition 34, the return words of w can be written as

$$r_1(w) = w\, m_1(w)\, t_2(w), \qquad r_2(w) = w\, t_2(w),$$

for some words $m_1(w)$ and $t_2(w)$. Let $Z := Y^{-1}(2)$ be the positions of the letter 2. If $t_2(w)$ is nonempty, then any letter in $t_2(w)$ occurs in $Y := \mathcal{T}(x_{\mathrm{F}})$ in positions which are just a shift $-|t_2(w)|, \dots, -1$ of Z, so each letter occurs according to a generalized Beatty sequence. The word $m_1(w)$ is never empty, and any letter in $m_1(w)$ occurs in $Y = \mathcal{T}(x_{\mathrm{F}})$ in positions which are a shift of a subsequence of Z (except, possibly, for the first occurrence, which then is in $r_0(w)$). This subsequence is obtained by replacing the distances ℓ_1 and ℓ_2 of ΔZ by $\ell_1 + \ell_2$ and ℓ_1. Moreover, these distances occur as the Fibonacci word x_{F} on the alphabet $\{\ell_1 + \ell_2, \ell_1\}$, because x_{F} is invariant under $0 \mapsto 01$, $1 \mapsto 0$. Thus each letter in $m_1(w)$ occurs according to a generalized Beatty sequence. All these $|t_2(w)| + |m_1(w)|$ sequences start at index 1. If we let the last $|r_0(w)|$ of these sequences start at index 0, then we have taken into account *all* elements of Y. This works, because of property (i) in Proposition 34. □

Here is an example where (SR0) is not satisfied.

Example 35. We take $\mathcal{T} : [10100 \mapsto 2]$, with image $\mathcal{T}(x_{\mathrm{F}}) = 01002100221002\dots$, so $Y = (5, 9, 10\dots)$. Here $r_0(w) = 0100$, $r_1(w) = 10100100$, $r_2(w) = 10100$. The positions of 0 are given by the sequence $(\mathcal{T}(x_{\mathrm{F}}))^{-1}(0) = 1, 3, 4, 7, 8\dots$, which can be written as a union of two generalized Beatty sequences, *except* that the position 1 from the first 0 in $\mathcal{T}(x_{\mathrm{F}})$ will not be in this union.

With (6) we can deduce an equivalent simple formulation of condition (SR0). If $w = \mu_1 s_k \mu_2$, then $r_0(w)$ equals $s_{k+1}\mu_1^{-1}$ with the first letter removed, and $r_1(w) = \mu_1 s_k s_{k+1}\mu_1^{-1}$, so

$$|w| = |\mu_1| + F_k + |\mu_2|, \qquad |r_0(w)| = F_{k+1} - |\mu_1| - 1, \qquad |r_1(w)| = F_{k+1} + F_k.$$

Filling this into condition (SR0) we obtain

$$|\mu_2| \leq 1. \tag{SR0$'$}$$

Using this condition, together with Theorem 6 in [Wen and Wen 1994], one can show that Theorem 33 does apply to at most 3 words w of length m, for all $m \geq 2$ (in fact, only 2, if m is not a Fibonacci number).

Acknowledgments

We thank the referee for useful comments. We also thank S. Levy, from the publisher's editorial staff, for several improvements to our text.

References

[Allauzen 1998] C. Allauzen, "Une caractérisation simple des nombres de Sturm", *J. Théor. Nombres Bordeaux* **10**:2 (1998), 237–241. MR Zbl

[Allouche and Shallit 2003] J.-P. Allouche and J. Shallit, *Automatic sequences: theory, applications, generalizations*, Cambridge University Press, 2003. MR Zbl

[Allouche et al. 2016] J.-P. Allouche, B. Cloitre, and V. Shevelev, "Beyond odious and evil", *Aequationes Math.* **90**:2 (2016), 341–353. MR Zbl

[Artstein-Avidan et al. 2008] S. Artstein-Avidan, A. S. Fraenkel, and V. T. Sós, "A two-parameter family of an extension of Beatty sequences", *Discrete Math.* **308**:20 (2008), 4578–4588. MR Zbl

[Ballot 2017] C. Ballot, "On functions expressible as words on a pair of Beatty sequences", *J. Integer Seq.* **20**:4 (2017), art. id. 17.4.2, 24 pp. MR Zbl

[Carlitz et al. 1972] L. Carlitz, R. Scoville, and V. E. Hoggatt, Jr., "Fibonacci representations", *Fibonacci Quart.* **10**:1 (1972), 1–28. Addendum in **10**:5 (1972), 527–530. MR Zbl

[Cassaigne 1998] J. Cassaigne, "Sequences with grouped factors", pp. 211–222 in *Developments in language theory* (Thessaloniki, 1997), 1998.

[Coven 1974/75] E. M. Coven, "Sequences with minimal block growth, II", *Math. Systems Theory* **8**:4 (1974/75), 376–382. MR Zbl

[Crisp et al. 1993] D. Crisp, W. Moran, A. Pollington, and P. Shiue, "Substitution invariant cutting sequences", *J. Théor. Nombres Bordeaux* **5**:1 (1993), 123–137. MR Zbl

[Dekking 2016] F. M. Dekking, "Morphisms, symbolic sequences, and their standard forms", *J. Integer Seq.* **19**:1 (2016), art. id. 16.1.1, 8 pp. MR Zbl

[Dekking 2018a] M. Dekking, "The Frobenius problem for homomorphic embeddings of languages into the integers", *Theoret. Comput. Sci.* **732** (2018), 73–79. MR Zbl

[Dekking 2018b] M. Dekking, "Substitution invariant Sturmian words and binary trees", *Integers* **18A** (2018), art. id. a7, 15 pp. Zbl

[Fraenkel 1977] A. S. Fraenkel, "Complementary systems of integers", *Amer. Math. Monthly* **84**:2 (1977), 114–115. MR Zbl

[Fraenkel 1994] A. S. Fraenkel, "Iterated floor function, algebraic numbers, discrete chaos, Beatty subsequences, semigroups", *Trans. Amer. Math. Soc.* **341**:2 (1994), 639–664. MR Zbl

[Fraenkel 2010a] A. S. Fraenkel, "Complementary iterated floor words and the Flora game", *SIAM J. Discrete Math.* **24**:2 (2010), 570–588. MR Zbl

[Fraenkel 2010b] A. S. Fraenkel, "From enmity to amity", *Amer. Math. Monthly* **117**:7 (2010), 646–648. MR Zbl

[Griffiths 2015] M. Griffiths, "On a matrix arising from a family of iterated self-compositions", *J. Integer Seq.* **18**:11 (2015), art. id. 15.11.8, 12 pp. Zbl

[Hildebrand et al. 2018] A. J. Hildebrand, J. Li, X. Li, and Y. Xie, "Almost Beatty partitions", preprint, 2018. arXiv

[Huang and Wen 2015] Y. Huang and Z. Wen, "The sequence of return words of the Fibonacci sequence", *Theoret. Comput. Sci.* **593** (2015), 106–116. MR Zbl

[Kimberling 2008] C. Kimberling, "Complementary equations and Wythoff sequences", *J. Integer Seq.* **11**:3 (2008), art. id. 08.3.3, 8 pp. MR Zbl

[Kimberling 2011] C. Kimberling, "Beatty sequences and Wythoff sequences, generalized", *Fibonacci Quart.* **49**:3 (2011), 195–200. MR Zbl

[Kimberling and Stolarsky 2016] C. Kimberling and K. B. Stolarsky, "Slow Beatty sequences, devious convergence, and partitional divergence", *Amer. Math. Monthly* **123**:3 (2016), 267–273. MR Zbl

[Lambek and Moser 1954] J. Lambek and L. Moser, "Inverse and complementary sequences of natural numbers", *Amer. Math. Monthly* **61** (1954), 454–458. MR Zbl

[Larsson et al. 2015] U. Larsson, N. A. McKay, R. J. Nowakowski, and A. A. Siegel, "Finding golden nuggets by reduction", preprint, 2015. arXiv

[Lind 1968] D. A. Lind, "The quadratic field $Q(\sqrt{5})$ and a certain Diophantine equation", *Fibonacci Quart.* **6**:3 (1968), 86–93. MR Zbl

[Lothaire 2002] M. Lothaire, *Algebraic combinatorics on words*, Encyclopedia of Mathematics and its Applications **90**, Cambridge University Press, 2002. MR Zbl

[Mercer 1978] A. M. Mercer, "Generalized Beatty sequences", *Internat. J. Math. Math. Sci.* **1**:4 (1978), 525–528. MR Zbl

[OEIS] N. J. A. Sloane et al., "The on-line encyclopedia of integer sequences", http://oeis.org/.

[Paul 1974/75] M. E. Paul, "Minimal symbolic flows having minimal block growth", *Math. Systems Theory* **8**:4 (1974/75), 309–315. MR Zbl

[Skolem 1957] T. Skolem, "On certain distributions of integers in pairs with given differences", *Math. Scand.* **5** (1957), 57–68. MR Zbl

[Tijdeman 1996] R. Tijdeman, "On complementary triples of Sturmian bisequences", *Indag. Math.* (*N.S.*) **7**:3 (1996), 419–424. MR Zbl

[Tijdeman 2000] R. Tijdeman, "Exact covers of balanced sequences and Fraenkel's conjecture", pp. 467–483 in *Algebraic number theory and Diophantine analysis* (Graz, 1998), edited by F. Halter-Koch and R. F. Tichy, de Gruyter, Berlin, 2000. MR Zbl

[Wen and Wen 1994] Z. X. Wen and Z. Y. Wen, "Some properties of the singular words of the Fibonacci word", *European J. Combin.* **15**:6 (1994), 587–598. MR Zbl

Received 28 Dec 2018. Revised 22 May 2019.

JEAN-PAUL ALLOUCHE:

jean-paul.allouche@imj-prg.fr
CNRS, IMJ-PRG, Sorbonne Université, Paris, France

F. MICHEL DEKKING:

f.m.dekking@math.tudelft.nl
Delft University of Technology, Faculty EEMCS, Delft, Netherlands

Moscow Journal of Combinatorics and Number Theory
Vol. 8, No. 4, 2019
dx.doi.org/10.2140/moscow.2019.8.343

Counting formulas for CM-types

Masanari Kida

We prove various counting formulas for CM-types of CM-fields and use them to construct infinite families of degenerate CM-types.

1. Introduction

Let M be a CM-field, which is, by definition, a totally imaginary quadratic extension of a totally real field M^+. Let ρ be a complex conjugation acting on M. Then M^+ is the maximal field fixed by ρ. Let Γ_M be the set of the complex embeddings of M into $\mathbb{C}$. If we denote by $2d$ the degree $[M:\mathbb{Q}]$, then we have $|\Gamma_M| = 2d$. A half set S of Γ_M is called a *CM-type* of M if it satisfies $\Gamma_M = S \sqcup S\rho$ (a disjoint union). Let L be the Galois closure of M over $\mathbb{Q}$ and $G = \mathrm{Gal}(L/\mathbb{Q})$ and $H = \mathrm{Gal}(L/M)$. We have a one-to-one correspondence between Γ_M and the right cosets $H\backslash G$ and ρ lifts to a central involution of G, which we also denote by ρ. We denote the set of the CM-types with respect to (G, H, ρ) by $\mathrm{CM}(G, H, \rho)$.

We define a family $\mathscr{H}$ of subgroups of G by

$$\mathscr{H} = \{H \leq G \mid \rho \notin H\}. \tag{1-1}$$

The fixed field of each $H \in \mathscr{H}$ is a CM-subfield of L; thus, we call such an H a *CM subgroup* of G. The set $\mathscr{H}$ of CM subgroups is a poset by inclusion. If $H \in \mathscr{H}$, we denote by $\pi_H : G \to H\backslash G$ the canonical surjection. Let $\widetilde{S} = \pi_H^{-1}(S)$ be the pullback of S to L. We define two subgroups of G by

$$s(S) = \{g \in G \mid g\widetilde{S} = \widetilde{S}\}, \tag{1-2}$$

$$r(S) = \{g \in G \mid \widetilde{S}g = \widetilde{S}\}. \tag{1-3}$$

It is easy to see that $s(S)$ and $r(S)$ are members of $\mathscr{H}$. A CM-type $S \in \mathrm{CM}(G, H, \rho)$ is called *simple* if $s(S) = H$. If $H' = r(S)$, then

$$S' = \pi_{H'}(\{x^{-1} \mid x \in \widetilde{S}\})$$

is a CM-type of (G, H', ρ) called the *reflex* CM-type of S. We call the group $H' = r(S)$ the *reflex subgroup* of S.

Two CM-types S_1 and S_2 in $\mathrm{CM}(G, H, \rho)$ are *conjugate* if there exists $g \in G$ such that $S_1 = S_2 g$. A conjugacy class of simple CM-types determines an isogeny class of complex abelian varieties with complex multiplication by an order of M.

MSC2010: 11G15, 14K22.
Keywords: CM-types, degenerate CM-types, Hodge conjecture.

The aim of this paper is to prove various counting formulas of the number of CM-types. Our fundamental formula (Theorem 2.4) counts the cardinality of the set

$$\{S \in \mathrm{CM}(G,1,\rho) \mid s(S) = H \text{ and } r(S) = K\}$$

for given $H, K \in \mathscr{H}$. Other counting formulas we shall prove are those of simple CM-types (Proposition 3.1) and CM-types of (G, H, ρ) with given reflex subgroup (Proposition 3.3) and conjugacy classes of CM-types (Theorem 4.1). These counting formulas enable us to construct degenerate CM-types (Section 5). In fact, we can construct infinite families of degenerate CM-types in Section 6 for nonabelian groups G. Infinite families of degenerate CM-types are previously known by [Greenberg 1980; Dodson 1984].

Throughout this paper, we will use the following purely group-theoretic setting. Let G be a finite group, and ρ a central involution of G fixed once for all. For $H \in \mathscr{H}$ (see (1-1)) we define a subposet $\mathscr{H}_{\geq}(H)$ of $\mathscr{H}$ by

$$\mathscr{H}_{\geq}(H) = \{K \in \mathscr{H} \mid K \geq H\}. \tag{1-4}$$

For $H_1, H_2 \in \mathscr{H}$ satisfying $H_1 \leq H_2$, the Möbius function μ on $\mathscr{H}$ is defined inductively by

$$\mu(H_1, H_1) = 1 \quad \text{and} \quad \mu(H_1, H_2) = -\sum_{H_1 \leq H < H_2} \mu(H_1, H_2). \tag{1-5}$$

2. Fundamental formula

In this section, we prove a counting formula of certain subsets of CM-types. The formula will play a fundamental role throughout the paper. The objects for counting are defined as follows. For $H, K \in \mathscr{H}$, we define

$$\mathscr{X}(H, K) = \{S \in \mathrm{CM}(G,1,\rho) \mid s(S) = H \text{ and } r(S) = K\}, \tag{2-1}$$

$$\mathscr{X}_{\geq}(H, K) = \{S \in \mathrm{CM}(G,1,\rho) \mid s(S) \geq H \text{ and } r(S) \geq K\}. \tag{2-2}$$

From these definitions, it readily follows that

$$\mathscr{X}_{\geq}(H, K) = \bigsqcup_{H_1 \in \mathscr{H}_{\geq}(H)} \bigsqcup_{K_1 \in \mathscr{H}_{\geq}(K)} \mathscr{X}(H_1, K_1). \tag{2-3}$$

The following function ε also plays an important role in the rest of this paper.

Definition 2.1. We define a function ε on $\mathscr{H} \times \mathscr{H}$ by

$$\varepsilon(H, K) = \begin{cases} 1 & \text{if } HK^g \not\ni \rho \text{ for all } g \in G, \\ 0 & \text{otherwise,} \end{cases}$$

where $H, K \in \mathscr{H}$ and K^g is the conjugate group gKg^{-1} of K.

We will need the following elementary properties of ε.

Lemma 2.2. *The function ε in Definition 2.1 satisfies the following properties*:

(i) $\varepsilon(H, 1) = \varepsilon(1, H) = 1$ *for all* $H \in \mathscr{H}$.

(ii) *If* $\varepsilon(H, K) = 0$, *then* $\varepsilon(H_1, K_1) = 0$ *for all* $H_1 \geq H$ *and all* $K_1 \geq K$.

(iii) $\varepsilon(H, K) = \varepsilon(K, H)$ *for all* $H, K \in \mathscr{H}$.

(iv) $\varepsilon(H, K) = \varepsilon(H, K^g)$ *for all* $H, K \in \mathscr{H}$ *and for all* $g \in G$.

Proof. (i) This follows from the fact that all $H \in \mathscr{H}$ do not contain ρ.

(ii) If $\rho \in HK^g$ for some g, then $\rho \in HK^g \leq H_1 K_1^g$ for all $H_1 \geq H$ and $K_1 \geq K$.

(iii) If $\rho \in KH^x$ for some $x \in G$, then $\rho = x^{-1}\rho^{-1}x \in HK^{x^{-1}}$ and vice versa.

(iv) If $\varepsilon(H, K) = 0$, then there exist $h \in H,\ k \in K,\ x \in G$ such that $\rho = hxkx^{-1}$. By writing $x = yg$, we have $\rho = hygkg^{-1}y^{-1} \in H(K^g)^y$. Thus we obtain $\varepsilon(H, K^g) = 0$. On the other hand, suppose that $\varepsilon(H, K) = 1$. If $\varepsilon(H, K^g) = 0$ for some $g \in G$, then there exists $y \in G$ such that $\rho \in H(K^g)^y = HK^{gy}$. This is a contradiction. Thus we have $\varepsilon(H, K^g) = 1$ for all $g \in G$. □

Lemma 2.3. *The following formula holds*:

$$|\mathscr{X}_{\geq}(H, K)| = \varepsilon(H, K) 2^{\frac{1}{2}|H\backslash G/K|} = \begin{cases} 0 & \textit{if } \varepsilon(K, H) = 0, \\ 2^{\frac{1}{2}|H\backslash G/K|} & \textit{otherwise}. \end{cases}$$

Proof. Let S be a CM-type in $\mathrm{CM}(G, 1, \rho)$. First note that if $s(S) \geq H$, then the natural projection π_H sends S to the right coset space $H\backslash G$ and S can be written as a union of right cosets of H:

$$S = Hg_1 \sqcup \cdots \sqcup Hg_s \quad (s = [G : H]).$$

Similarly, if $r(S) \geq K$, then a natural map sends S to the left cosets space G/K and, thus, gives a left coset decomposition of S:

$$S = t_1 K \sqcup \cdots \sqcup t_u K \quad (u = [G : K]).$$

Therefore if $S \in \mathscr{X}_{\geq}(N, K)$, then we have a double coset decomposition

$$S = Hx_1 K \sqcup \cdots \sqcup Hx_r K.$$

If $\varepsilon(H, K) = 1$, then both x_i and ρx_i cannot belong to a same double coset $Hx_i K$ simultaneously. Suppose to the contrary that $Hx_i K = H\rho x_i K$. This equality means that there exist $h \in H$ and $k \in K$ such that $x_i = h\rho x_i k$. Since ρ is central in G, this, in turn, implies $\rho = hx_i k x_i^{-1} \in HK^{x_i}$. This is a contradiction. Thus if $\varepsilon(H, K) = 1$, then x_i and ρx_i belong to different double cosets and we have a double coset decomposition of G of the form

$$G = Hx_1 K \sqcup \cdots \sqcup Hx_r K \sqcup H\rho x_1 K \sqcup \cdots \sqcup H\rho x_r K \tag{2-4}$$

and hence we have $2r = |H\backslash G/K|$.

Conversely, if we have the double coset decomposition (2-4), then by choosing one double coset from each pair of cosets $(Hx_i K, H\rho x_i K)$, we can form a CM-type S such that $s(S) \geq N$ and $r(S) \geq K$.

Hence under the assumption $\varepsilon(H, K) = 1$, we have established a one-to-one correspondence between $\mathscr{X}_{\geq}(H, K)$ and the pairwise choice from the double coset decomposition of the form (2-4). We conclude that

$$|\mathscr{X}_{\geq}(N, K)| = 2^{\frac{1}{2}|H\backslash G/K|}$$

if $\varepsilon(H, K) = 1$. On the other hand, if $\varepsilon(H, K) = 0$, then the double coset decomposition of the form (2-4) is obviously impossible. Thus there is no CM-type satisfying the conditions. □

Now we state and prove our fundamental formula.

Theorem 2.4. *The number of elements in $\mathscr{X}(H,K)$ defined by (2-1) is given by*

$$|\mathscr{X}(H,K)| = \sum_{H_1\in\mathscr{H}_{\geq}(H)} \sum_{K_1\in\mathscr{H}_{\geq}(K)} \varepsilon(H_1,K_1)\mu(H,H_1)\mu(K,K_1)2^{\frac{1}{2}|H_1\backslash G/K_1|}.$$

Proof. We consider a product poset $\mathscr{H}\times\mathscr{H}$ whose partial order is defined by

$$(H_1,K_1)\geq(H_2,K_2) \quad\Longleftrightarrow\quad H_1\geq H_2 \text{ and } K_1\geq K_2.$$

Then the identity (2-3) can be rewritten as

$$\mathscr{X}_{\geq}(H,K) = \bigsqcup_{(H_1,K_1)\geq(H,K)} \mathscr{X}(H_1,K_1).$$

The Möbius inversion formula [Rota 1964, Proposition 3] on $\mathscr{H}\times\mathscr{H}$ implies

$$|\mathscr{X}(H,K)| = \sum_{(H_1,K_1)\geq(H,K)} \mu((H,K),(H_1,K_1))|\mathscr{X}_{\geq}(H_1,K_1)|.$$

Since the Möbius function on the product poset is nothing but a product of corresponding Möbius functions [Rota 1964, Proposition 5], we have

$$|\mathscr{X}(H,K)| = \sum_{H_1\in\mathscr{H}_{\geq}(H)} \sum_{K_1\in\mathscr{H}_{\geq}(K)} \mu(H,H_1)\mu(K,K_1)|\mathscr{X}_{\geq}(H_1,K_1)|.$$

By Lemma 2.2(ii), neglecting the terms with $\varepsilon(H_1,K_1)=0$ does not affect the Möbius function. □

By the symmetry of the formula of Theorem 2.4 and that of ε (Lemma 2.2(iii)), we have:

Corollary 2.5. *For any $H,K\in\mathscr{H}$, we have $|\mathscr{X}(H,K)|=|\mathscr{X}(K,H)|$.*

The following corollaries help us to conclude $\mathscr{X}(H,K)=\varnothing$ when H is a normal subgroup of G.

Corollary 2.6. *Let $H,K\in\mathscr{H}$. Suppose that H is a normal subgroup of G:*

(i) *If $H\geq K$, then $\mathscr{X}(H,K)\neq\varnothing$ if and only if $H=K$.*

(ii) *If $HK\gneq K$, then $\mathscr{X}(H,K)=\varnothing$.*

Proof. If H is normal in G, then the double coset decomposition (2-4) agrees with the left coset decomposition by HK. If $H\geq K$ as in (i), then $HK=H$. Hence the result follows. Also if $HK\gneq K$ as in (ii), then the reflex subgroup must be strictly larger than K and $\mathscr{X}(H,K)=\varnothing$. □

Corollary 2.7. *Let $H,K\in\mathscr{H}$. Suppose that both H and K are normal subgroups of G. If $H\neq K$, then*

$$\mathscr{X}(H,K)=\varnothing.$$

In particular, if all CM subgroups of G are normal, then the matrix $(|\mathscr{X}(H,K)|)_{H,K\in\mathscr{H}}$ is diagonal.

Proof. If $H\neq K$, then $HK\gneq K$ holds. Hence by Corollary 2.6(ii), we have $\mathscr{X}(H,K)=\varnothing$. □

If, in particular, all the subgroups of G are normal (such finite groups are called Dedekind groups), then the second assertion of Corollary 2.7 can apply. Dedekind classified such groups: they are of the form $Q_8\times A$ where A is an abelian group whose 2-Sylow subgroup is elementary. Other than Dedekind groups, nonabelian groups whose CM subgroups with respect to some central involution are all normal include the generalized quaternion groups Q_{2^n} and many others.

Corollary 2.8. *For all $H, K \in \mathscr{H}$ and all $x, y \in G$, we have*

$$|\mathscr{X}(H, K)| = |\mathscr{X}(H^x, K^y)|.$$

Proof. If $\varepsilon(H, K) = 1$, then there is a double coset decomposition of G by H and K given by (2-4). By Lemma 2.2(iv), there also exists a double coset decomposition by H and K^x for all $x \in G$. In fact, it is given by

$$G = Gx^{-1} = Hx_1x^{-1}K^x \sqcup \cdots \sqcup Hx_rx^{-1}K^x \sqcup H\rho x_1x^{-1}K^x \sqcup \cdots \sqcup H\rho x_rx^{-1}K^x.$$

In particular, we have $|\mathscr{X}_{\geq}(H, K)| = |\mathscr{X}_{\geq}(H, K^x)|$. The corollary now follows from Corollary 2.5. □

3. Simple CM-types and reflex CM-types

Let $H \in \mathscr{H}$. In this section, we enumerate simple CM-types in $\mathrm{CM}(G, H, \rho)$ and CM-types in $\mathrm{CM}(G, H, \rho)$ whose reflex subgroup coincides with given $K \in \mathscr{H}$. Although these sets are unions of some $\mathscr{X}(H', K')$'s in Section 2, we can obtain simpler formulas than that can be derived from Theorem 2.4. These formulas will be required to compute the number of conjugacy classes in Section 4.

For that purpose, we define

$$\mathscr{S}(H) = \{S \in \mathrm{CM}(G, 1, \rho) \mid s(S) = H\}, \tag{3-1}$$

$$\mathscr{S}_{\geq}(H) = \{S \in \mathrm{CM}(G, 1, \rho) \mid s(S) \geq H\}, \tag{3-2}$$

$$\mathscr{R}(H) = \{S \in \mathrm{CM}(G, 1, \rho) \mid r(S) = H\}, \tag{3-3}$$

$$\mathscr{R}_{\geq}(H) = \{S \in \mathrm{CM}(G, 1, \rho) \mid r(S) \geq H\}, \tag{3-4}$$

where $s(S)$ and $r(S)$ are defined by (1-2) and (1-3), respectively.

The set $\mathscr{S}(H)$ consists of the pullbacks of the simple CM-types of $\mathrm{CM}(G, H, \rho)$, while $\mathscr{S}_{\geq}(H)$ is the set of the pullbacks of $\mathrm{CM}(G, H, \rho)$.

Proposition 3.1. *The number of the simple CM-types in* $\mathrm{CM}(G, H, \rho)$ *is given by*

$$|\mathscr{S}(H)| = \sum_{N \in \mathscr{H}_{\geq}(H)} \mu(H, N) 2^{\frac{1}{2}|N \backslash G|}.$$

Proof. The following equality obviously holds by definition:

$$\mathscr{S}_{\geq}(H) = \bigsqcup_{N \in \mathscr{H}_{\geq}(H)} \mathscr{S}(N).$$

The cardinality of $\mathscr{S}_{\geq}(H)$ is given by

$$|\mathscr{S}_{\geq}(H)| = |\pi_H^{-1}(\mathrm{CM}(G, H, \rho))| = |\mathrm{CM}(G, H, \rho)| = 2^{\frac{1}{2}|H \backslash G|}.$$

Hence simple Möbius inversion implies our result. □

Corollary 3.2. *The cardinality of $\mathscr{R}(H)$ is the same as that of $\mathscr{S}(H)$:*

$$|\mathscr{R}(H)| = \sum_{N \in \mathscr{H}_{\geq}(H)} \mu(H, N) 2^{\frac{1}{2}|N \backslash G|}.$$

Proof. By the definitions (3-1) and (3-3), it is easy to see that

$$\mathscr{S}(H) = \bigsqcup_{K \in \mathscr{H}} \mathscr{X}(H, K) \quad \text{and} \quad \mathscr{R}(H) = \bigsqcup_{K \in \mathscr{H}} \mathscr{X}(K, H).$$

From Corollary 2.5, the result follows. □

Let $H, K \in \mathscr{H}$. We next count the number of CM-types in $\mathrm{CM}(G, H, \rho)$ whose reflex subgroup is K, namely the cardinality of the set

$$\pi_H(\mathscr{S}_{\geq}(H) \cap \mathscr{R}(K)) = \{S \in \mathrm{CM}(G, H, \rho) \mid r(S) = K\}.$$

Proposition 3.3. *We have the following formula*:

$$|\pi_H(\mathscr{S}_{\geq}(H) \cap \mathscr{R}(K))| = \sum_{N \in \mathscr{H}_{\geq}(K)} \varepsilon(H, N)\mu(K, N)2^{\frac{1}{2}|H\backslash G/N|}.$$

Proof. In the right-hand side of (2-3), we see

$$\begin{aligned}\bigsqcup_{H_1 \in \mathscr{H}_{\geq}(H)} \mathscr{X}(H_1, K_1) &= \{S \in \mathrm{CM}(G, 1, \rho) \mid s(S) \geq H \text{ and } r(S) = K_1\} \\ &= \mathscr{S}_{\geq}(H) \cap \mathscr{R}(K_1)\end{aligned}$$

and thus

$$\mathscr{X}_{\geq}(H, K) = \bigsqcup_{K_1 \in \mathscr{H}_{\geq}(K)} \mathscr{S}_{\geq}(H) \cap \mathscr{R}(K_1).$$

Möbius inversion implies

$$|\mathscr{S}_{\geq}(H) \cap \mathscr{R}(K)| = \sum_{K_1 \in \mathscr{H}_{\geq}(K)} \mu(K, K_1)|\mathscr{X}_{\geq}(H, K_1)|.$$

By Lemma 2.2(ii), the function $\varepsilon(H, N)$ is compatible with the poset structure of $\mathscr{H}_{\geq}(K)$, and we obtain the proposition using Lemma 2.3. □

Corollary 3.4. *We have* $|\mathscr{S}(H) \cap \mathscr{R}_{\geq}(K)| = |\mathscr{S}_{\geq}(K) \cap \mathscr{R}(H)|$.

Proof. By Corollary 2.5, we see that

$$|\mathscr{S}(H) \cap \mathscr{R}_{\geq}(K)| = \sum_{K_1 \in \mathscr{H}_{\geq}(K)} |\mathscr{X}(H, K_1)| = \sum_{K_1 \in \mathscr{H}_{\geq}(K)} |\mathscr{X}(K_1, H)| = |\mathscr{S}_{\geq}(K) \cap \mathscr{R}(H)|. \quad \square$$

For convenience, we summarize the counting formulas obtained up to this section. We arrange the members of $\mathscr{H}$ in a line so that $H_i > H_j$ implies $i > j$ and we form a matrix $X = (|\mathscr{X}(H, K)|)_{H,K \in \mathscr{H}}$. The counting formulas are summarized as follows:

each entry $\lvert\mathscr{X}(H, K)\rvert$	Theorem 2.4
a row sum $\lvert\mathscr{S}(H)\rvert$	Proposition 3.1
a column sum $\lvert\mathscr{R}(K)\rvert$	Corollary 3.2
a row subsum $\lvert\mathscr{S}(H) \cap \mathscr{R}_{\geq}(K)\rvert$	Corollary 3.4
a column subsum $\lvert\mathscr{S}_{\geq}(H) \cap \mathscr{R}(K)\rvert$	Proposition 3.3
a submatrix sum $\lvert\mathscr{X}_{\geq}(H, K)\rvert$	Lemma 2.3

4. Number of conjugacy classes

In this section, we prove a counting formula for the conjugacy classes in $\mathrm{CM}(G, H, \rho)$.

Recall that two CM-types $S, S' \in \mathrm{CM}(G, H, \rho)$ are conjugate if there exists $g \in G$ satisfying $S' = Sg$. Therefore the number of conjugacy classes is the number of the orbits under this group action.

Theorem 4.1. *Let G be a finite group and ρ a central involution of G and H a subgroup of G not containing ρ. The number $c(G, H, \rho)$ of the conjugacy classes in $\mathrm{CM}(G, H, \rho)$ is given by*

$$\begin{aligned} c(G, H, \rho) &= \frac{1}{|G|} \sum_{K \in \mathscr{H}} |K| |\mathscr{S}_{\geq}(H) \cap \mathscr{R}(K)| \\ &= \frac{1}{|G|} \sum_{K \in \mathscr{H}} |K| \sum_{N \in \mathscr{H}_{\geq}(K)} \varepsilon(H, N) \mu(K, N)\, 2^{\frac{1}{2}|H \backslash G/N|}. \end{aligned}$$

To prove Theorem 4.1, we need the following proposition.

Proposition 4.2. *For $H, K \in \mathscr{H}$, the number of conjugacy classes in $\mathscr{S}_{\geq}(H) \cap \mathscr{R}(K)$ is*

$$\frac{|K|}{|G|} |\mathscr{S}_{\geq}(H) \cap \mathscr{R}(K)|.$$

Proof. If $g \in G$ and $S \in \mathrm{CM}(G, H, \rho)$, then $Sg = S$ holds if and only if $g \in r(S)$. Hence the g-invariant subset of $(\mathscr{S}_{\geq}(H) \cap \mathscr{R}(K))$ is given by

$$\begin{aligned} (\mathscr{S}_{\geq}(H) \cap \mathscr{R}(K))^g &= \{S \in \mathscr{S}_{\geq}(H) \cap \mathscr{R}(K) \mid S = Sg\} \\ &= \begin{cases} \mathscr{S}_{\geq}(H) \cap \mathscr{R}(K) & \text{if } g \in K, \\ \varnothing & \text{otherwise.} \end{cases} \end{aligned}$$

Thus we obtain

$$|(\mathscr{S}_{\geq}(H) \cap \mathscr{R}(K))^g| = \mathrm{ch}_K(g) |\mathscr{S}_{\geq}(H) \cap \mathscr{R}(K)|,$$

where

$$\mathrm{ch}_K(g) = \begin{cases} 1 & \text{if } g \in K, \\ 0 & \text{otherwise} \end{cases}$$

is the characteristic function of K. From the lemma of Burnside and Frobenius [Aigner 2007, Lemma 6.2] it follows that the number of the orbits is then given by

$$\frac{1}{|G|} \sum_{g \in K} \mathrm{ch}_K(g)\, |\mathscr{S}_{\geq}(H) \cap \mathscr{R}(K)| = \frac{|K|}{|G|} |\mathscr{S}_{\geq}(H) \cap \mathscr{R}(K)|. \qquad \square$$

Now we can prove the counting formula for the conjugacy classes.

Proof of Theorem 4.1. If two CM-types S and S' are conjugate, then we have $r(S) = r(S')$ by the definition of the conjugacy. Thus the decomposition $\mathrm{CM}(G, H, \rho) = \bigsqcup_{K \in \mathscr{H}} \mathscr{S}_{\geq}(H) \cap \mathscr{R}(K)$ is stable under this G-action. Hence we have

$$\begin{aligned} c(G,H,\rho) &= \sum_{K\in\mathscr{H}} |\text{the conjugacy classes in } \mathscr{S}_{\geq}(H)\cap\mathscr{R}(K)| \\ &= \sum_{K\in\mathscr{H}} \frac{|K|}{|G|}\,|\mathscr{S}_{\geq}(H)\cap\mathscr{R}(K)| \end{aligned}$$

by Proposition 4.2. This is the first equality of the theorem. The second equality follows readily from Proposition 3.3. This completes the proof of Theorem 4.1. □

Remark 4.3. By Theorem 4.1, we have

$$\begin{aligned} |\mathrm{CM}(G,H,\rho)| &= \sum_{K\in\mathscr{H}} |\mathscr{S}_{\geq}(H)\cap\mathscr{R}(K)| \\ &= \sum_{K\in\mathscr{H}} \frac{|G|}{|K|}\,|\text{the conjugacy classes in } \mathscr{S}_{\geq}(H)\cap\mathscr{R}(K)|. \end{aligned}$$

For each $K\in\mathscr{H}$, we see $[L^K:\mathbb{Q}]=|G|/|K|$, where L^K is the reflex field. Hence the sum of $[L^K:\mathbb{Q}]$ over a representative of the conjugacy classes of $\mathrm{CM}(G,H,\rho)$ is $2^{\frac{1}{2}|H\backslash G|}$. This fact was previously noticed by Dodson [1984, p. 5] and Oishi-Tomiyasu [2010, Lemma 1.4].

5. Construction of degenerate CM-types

For a simple CM-type $S\in\mathrm{CM}(G,H,\rho)$, let $H'=r(S)$ be the reflex subgroup. We define a linear map

$$\Phi_S:\mathbb{Z}[H\backslash G]\to\mathbb{Z}[H'\backslash G]$$

by $x\mapsto\sum_{H'\sigma\in S'}H'\sigma x$, where S' is the reflex CM-type of S. Here we understand that $\mathbb{Z}[H\backslash G]$ is a free left $\mathbb{Z}$-module on $H\backslash G$ and $\mathbb{Z}[H'\backslash G]$ is that on $H'\backslash G$ on which $H\backslash G$ acts from the right. Since the elements of the form $(Hx+Hx\rho)-(Hy+Hy\rho)$ with $x,y\in G$ are contained in the kernel of Φ_S, the rank of Φ_S is less than or equal to $\frac{1}{2}|H\backslash G|+1$. The CM-type S is called *nondegenerate* if the rank is maximal, that is, if $\operatorname{rank}\Phi_S=\frac{1}{2}|H\backslash G|+1$ holds; otherwise it is called *degenerate*. If S is nondegenerate, then the Hodge conjecture is true for CM abelian varieties with CM-type S, whereas, if S is degenerate, then exceptional Hodge cycles exist on a self-product of the CM abelian variety. Hence it is interesting to know how to construct degenerate CM-types. One easy way to construct degenerate simple CM-types S of (G,H,ρ) is to construct CM-types S satisfying $|H'|>|H|$, since we know the rank of Φ_S is less than or equal to $\min\{\frac{1}{2}|H\backslash G|+1,\frac{1}{2}|H'\backslash G|+1\}$ by [Ribet 1980, (3.5)].

We have the following criteria for such CM-types to exist. The first criterion is obvious from the definition of $\mathscr{X}(H,K)$ and Corollary 2.5.

Proposition 5.1. *If $\mathscr{X}(H,K)\neq\varnothing$ for some $H,K\in\mathscr{H}$ such that $|H|\neq|K|$, then there exists a degenerate CM-type in* $\mathrm{CM}(G,H,\rho)$ *or* $\mathrm{CM}(G,K,\rho)$.

Although we have a formula for $\mathscr{X}(H,K)$, it is not immediate how to determine whether $\mathscr{X}(H,K)=\varnothing$ or not. Indeed, there is an example of H such that $\mathscr{X}(H,K)=\varnothing$ for all $K\in\mathscr{H}$ (see Example 6.2).

The following proposition is sometimes useful.

Proposition 5.2. *Let $H \in \mathscr{H}$. Assume that $\varepsilon(H, K) = 0$ for all $K \in \mathscr{H}$ such that $|K| = |H|$. If there exists a simple CM-type in* $\mathrm{CM}(G, H, \rho)$, *then there exists a degenerate CM-type in* $\mathrm{CM}(G, H, \rho)$.

Proof. By the assumptions and Lemma 2.3, it is impossible to have a double coset decomposition by H and K like (2-4). If $H = 1$, then the assumption apparently does not hold and, hence, we may assume $H \neq 1$. Then there exists $N \in \mathscr{H}$ such that $|N| \neq |H|$ and $HN^x \not\ni \rho$ for all $x \in G$. For example, we can take $N = 1$. The double coset decomposition of G by H and N is now of the form (2-4). By the assumption of the proposition, the order of $r(S)$ of a simple CM-type S with respect to (G, H, ρ) is different from $|H|$. If $|r(S)| < |H|$, then replacing H by $r(S)$, we obtain a CM-type satisfying $|r(S)| > |H|$. □

In the next section, we will construct infinite families of pairs of finite groups (G, H) satisfying the conditions of Propositions 5.1 and 5.2.

6. Examples

In this section, we use our theorems to give several examples.

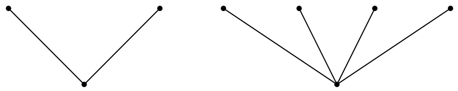
The following lemma is useful in explicit computation and interesting in its own right.

Lemma 6.1. *If $H \in \mathscr{H}$ is a normal subgroup of G such that the quotient G/H is isomorphic to either the direct product $C_2 \times C_2$ of cyclic groups of order* 2 *or the dihedral group D_4 of order* 8, *then we have*

$$\mathscr{S}(H) = \mathscr{R}(H) = \varnothing.$$

In particular, every CM abelian variety with CM-types in $\mathrm{CM}(G, H, \rho)$ *splits.*

Proof. If H is a normal subgroup, then there is a one-to-one correspondence between $\mathrm{CM}(G, H, \rho)$ and $\mathrm{CM}(G/H, 1, \rho H)$, where the latter ρ is the image of ρ under the natural projection. Therefore it suffices to show that $\mathscr{S}(1) = \mathscr{H}(1) = \varnothing$ for $G = C_2 \times C_2$ or $G = D_4$. For these two groups, the CM subgroups are of order 1 or 2 and the Hasse diagrams of $\mathscr{H}$ are



respectively. By Proposition 3.1, for $G = C_2 \times C_2$ we have $|\mathscr{S}(1)| = 2^2 - 2 \cdot 2 = 0$ and for $G = D_4$ we have $|\mathscr{S}(1)| = 2^4 - 4 \cdot 2^2 = 0$ as desired. The claim for $\mathscr{R}(H)$ follows from Corollary 3.2. □

Schappacher [1977] proved that the converse of Lemma 6.1 also holds.

Example 6.2 (cyclic group C_{2p}). Let p be an odd prime number and $G = C_{2p}$ a cyclic group of order $2p$ generated by x. The element $\rho = x^p$ is a unique central involution in G and we have $\mathscr{H} = \{\langle x^2\rangle, 1\}$. Since all subgroups of G are normal in G, it follows from Corollary 2.7 that if $H \neq K$, then $\mathscr{X}(H, K) = \varnothing$. We have to compute only $|\mathscr{X}(1, 1)|$ and $|\mathscr{X}(\langle x^2\rangle, \langle x^2\rangle)|$. The Möbius function on $\mathscr{H}$ is computed as

$$\mu(1, 1) = 1, \quad \mu(1, \langle x^2\rangle) = -1 \quad \text{and} \quad \mu(\langle x^2\rangle, \langle x^2\rangle) = 0.$$

By Theorem 2.4, we have

$$\begin{aligned} |\mathscr{X}(1, 1)| &= \mu(1, 1)^2 2^p + 2\mu(1, 1)\mu(1, \langle x^2\rangle)2 + \mu(1, \langle x^2\rangle)^2 2 = 2^p - 2, \\ |\mathscr{X}(\langle x^2\rangle, \langle x^2\rangle)| &= \mu(\langle x^2\rangle, \langle x^2\rangle)2 = 2. \end{aligned}$$

The number of conjugacy classes can be computed by Theorem 4.1 and, in this case, it is convenient to use

$$c(G,H,\rho)=\frac{1}{|G|}\sum_{K\in\mathscr{H}}|K||\mathscr{S}_{\geq}(H)\cap\mathscr{R}(K)|=\frac{1}{|G|}\sum_{K\in\mathscr{H}}\sum_{H_1\in\mathscr{H}_{\geq}(H)}|K||\mathscr{X}(H_1,K)|,$$

and we have

$$c(G,\langle x^2\rangle,\rho)=1,\quad c(G,1,\rho)=\frac{2^{p-1}-1}{p}+1.$$

The first term of the right-hand side of the second expression is an integer by Fermat's theorem. In particular, the situation discussed in Section 5 does not occur.

CM-fields over $\mathbb{Q}$ with Galois group C_{2p} can be constructed easily as follows: We choose a prime number q such that $p \parallel (q-1)$. Then the q-th cyclotomic field contains a unique totally real cyclic extension M of degree p over $\mathbb{Q}$. The composite field of M with an imaginary quadratic field gives a desired field.

Example 6.3 (dihedral group D_{2p}). Again let p be an odd prime number. We consider the dihedral group $G=D_{2p}$ of order $4p$, which has a presentation

$$D_{2p}=\langle s,t\mid s^2=1,\ t^{2p}=1,\ sts=t^{2p-1}\rangle.$$

A unique central involution in D_{2p} is $\rho=t^p$. The members of $\mathscr{H}$ are: two nonconjugate subgroups $H_1=\langle st\rangle$, ${H_1}'=\langle st^2\rangle$ of order 2 whose lengths are p and two normal subgroups $H_2=\langle st,t^2\rangle$, ${H_2}'=\langle st^2,t^2\rangle$ of order $2p$ and one normal subgroup $H_2\cap {H_2}'=\langle t^2\rangle$ of order p. The conjugates of H_1 and ${H_1}'$ are, respectively, $H_1^{t^i}$ and ${H_1}'^{t^i}$ $(i=0,1,\dots,p-1)$. The Hasse diagram of $\mathscr{H}$ modulo the conjugacy is

$$\begin{tikzcd}
H_2 \arrow[dr, dash] \arrow[dd, dash] & & {H_2}' \arrow[dl, dash] \arrow[dd, dash] \\
& H_2\cap {H_2}' \arrow[dd, dash] & \\
H_1 \arrow[dr, dash] & & {H_1}' \arrow[dl, dash] \\
& 1 &
\end{tikzcd}$$

Since $H_1({H_1}')^{t^{\frac{1}{2}(p+1)}}\ni st\cdot t^{\frac{1}{2}(p+1)}st^2t^{-\frac{1}{2}(p+1)}=t^p=\rho$, we have the following table of $\varepsilon=\varepsilon_{\mathscr{H}}$:

$H\backslash K$	1	H_1	H_1'	$H_2\cap H_2'$	H_2	H_2'
1	1	1	1	1	1	1
H_1	1	1	0	1	1	0
H_1'	1	0	1	1	0	1
$H_2\cap H_2'$	1	1	1	1	1	1
H_2	1	1	0	1	1	0
H_2'	1	0	1	1	0	1

The Möbius function $\mu(H,K)$ on $\mathscr{H}$ can be computed by the definition (1-5) by noting that both H_1 and H_2 have p conjugate groups:

$H\backslash K$	1	H_1	H_1'	$H_2\cap H_2'$	H_2	H_2'
1	1	-1	-1	-1	p	p
H_1	0	1	0	0	-1	0
H_1'	0	0	1	0	0	-1
$H_2\cap H_2'$	0	0	0	1	-1	-1
H_2	0	0	0	0	1	0
H_2'	0	0	0	0	0	1

If $N\in\mathscr{H}$ is normal in G, then $|H_1\backslash G/N|=|H_1N\backslash G|$, and hence, we have only to compute a double coset decomposition of G by H_1 and H_1:

$$G=H_1H_1\sqcup H_1tH_1\sqcup\cdots\sqcup H_1t^pH_1.$$

This yields $|H_1\backslash G/H_1|=p+1$.

To compute $|\mathscr{X}(1,H_1)|$, it is convenient to use Proposition 3.3. Since H_2 is normal in G, we have $\mathscr{X}(1,H_2)=\varnothing$ by Corollary 2.6. Therefore we obtain

$$\mathscr{S}(1)\cap\mathscr{R}_{\geq}(H_1)=\bigsqcup_{K\in\mathscr{H}_{\geq}(H_1)}\mathscr{X}(1,K)=\mathscr{X}(1,H_1)\sqcup\mathscr{X}(1,H_2)=\mathscr{X}(1,H_1).$$

Using Corollary 3.4 and Proposition 3.3, we compute

$$\begin{aligned}
&|\mathscr{X}(1,H_1)|\\
&\quad=|\mathscr{S}(1)\cap\mathscr{R}_{\geq}(H_1)|=|\mathscr{S}_{\geq}(H_1)\cap\mathscr{R}(1)|\\
&\quad=\sum_{N\in\mathscr{H}_{\geq}(1)}\varepsilon(H_1,N)\mu(1,N)2^{\frac{1}{2}|H_1\backslash G/N|}\\
&\quad=\mu(1,1)2^{\frac{1}{2}|H_1\backslash G|}+p\mu(1,H_1)2^{\frac{1}{2}|H_1\backslash G/H_1|}+\mu(1,H_2\cap H_2')2^{\frac{1}{2}|H_1\backslash G/H_2\cap H_2'|}+\mu(1,H_2)2^{\frac{1}{2}|H_1\backslash G/H_2|}\\
&\quad=2^p-p2^{\frac{1}{2}(p+1)}+2p-2.
\end{aligned}$$

This quantity is positive if $p\geq 7$. By Proposition 5.1, CM-types contained in this set are simple and degenerate. It is interesting to note that

$$\lim_{p\to\infty}\frac{|\mathscr{X}(1,H_1)|}{|\mathrm{CM}(D_{2p},1,\rho)|}=0.$$

We also compute $\mathscr{X}(H_2\cap H_2',K)$ for all $K\in\mathscr{H}$. Let $H=H_2\cap H_2'$ for short. Since H is normal in G, we have $\mathscr{X}(H,K)=\varnothing$ for $K=1,H_1,H_2$ by Corollary 2.6 and this also holds for $K=H_2$ and H_2' if we combine Corollaries 2.5 and 2.6. Thus only $\mathscr{X}(H,H)$ remains. On the other hand, since $G/H\cong D_4$, we have $\mathscr{S}(H)=\varnothing$ by Lemma 6.1 and, thus, $\mathscr{X}(H,H)$ must be empty. Hence $\mathscr{X}(H_2\cap H_2',K)=\varnothing$ holds for all $K\in\mathscr{H}$.

Example 6.4 (semidirect product $C_{2^k}\rtimes C_2$). Let k be an integer greater than or equal to 3. In this example, we consider semidirect products $C_{2^k}\rtimes C_2$, where $C_2=\langle s\rangle$ acts on $C_{2^k}=\langle t\rangle$ by $sts=t^u$. In this situation, u is one of -1, $2^{k-1} \pm 1$. If $u = -1$, then $C_{2^k} \rtimes C_2 \simeq D_{2^k}$ and if $u = 2^{k-1}+1$, then the group is isomorphic to the semidihedral group $SD_{2^{k+1}}$ and if $u = 2^{k-1}-1$, then the group is called the modular maximal-cyclic group and we denote it by $M_{2^{k+1}}$. A unique central involution of these groups is $\rho = t^{2^{k-1}}$. The CM subgroup posets are

$$\begin{aligned}\mathscr{H}(D_{2^k}) &= \{H_1 = \langle st\rangle,\ H_2 = \langle st^2\rangle\},\\ \mathscr{H}(SD_{2^{k+1}}) &= \{H_3 = \langle s\rangle\},\\ \mathscr{H}(M_{2^{k+1}}) &= \{H_4 = \langle s\rangle\}.\end{aligned}$$

The lengths of H_1, H_2, and H_3 are 2^{k-1} and that of H_4 is 2. Since $H_1 H_1^{t^{2^{k-2}}} \ni stt^{2^{k-2}}stt^{-2^{k-2}} = \rho$, we have $\varepsilon(H_1, H_1) = 0$. Similarly $H_2 H_2^{t^{2^{k-2}}}$, $H_3 H_3^{t^{-1}}$, and $H_4 H_4^{t}$ contain ρ and we conclude

$$\varepsilon(H_i, H_i) = 0 \quad (i = 1, 2, 3, 4). \tag{6-1}$$

Hence, in particular, $SD_{2^{k+1}}$ and $M_{2^{k+1}}$ satisfy the assumption in Proposition 5.2. On the other hand, we have

$$D_{2^k} = \bigsqcup_{i=1}^{2^{k-2}} (H_1 t^i H_2 \sqcup H_1 t_i \rho H_2) \tag{6-2}$$

and $\varepsilon(H_1, H_2) = 1$.

We compute $|\mathscr{X}(1, H_i)|$ by computing $|\mathscr{S}_{\geq}(H_i) \cap \mathscr{R}(1)|$ as in Example 6.3.

For H_1 and H_2, using (6-1) we have

$$\begin{aligned}|\mathscr{X}(1, H_i)| &= \sum_{N \in \mathscr{H}_{\geq}(1)} \varepsilon(H_i, N)\mu(1, N) 2^{\frac{1}{2}|H_i \backslash D_{2^k}/N|}\\ &= \varepsilon(H_i, 1)\mu(1, 1)2^{\frac{1}{2}|H_i\backslash G|} + \sum_x \varepsilon(H_i, H_j^x)\mu(1, H_j^x) 2^{\frac{1}{2}|H_i\backslash D_{2^k}/H_j|},\end{aligned}$$

where $i, j \in \{1, 2\}$ and $i \neq j$ and the last summation is taken over a transversal of $N_G(H_j)\backslash D_{2^k}$. From Lemma 2.2(iv) and (6-2) it follows

$$|\mathscr{X}(1, H_i)| = 2^{2^{k-1}} - 2^{k-1}2^{2^{k-2}} = 2^{2^{k-1}} - 2^{2^{k-2}+k+1} \quad (i = 1, 2),$$

which is positive if k is greater than 3.

For H_3 and H_4, computation is simpler. In fact, we have

$$|\mathscr{X}(1, H_i)| = \varepsilon(H_i, 1)\mu(1, 1)2^{\frac{1}{2}|H_i\backslash G|} = 2^{2^{k-1}},$$

where G is either $SD_{2^{k+1}}$ or $M_{2^{k+1}}$.

Example 6.5 (wreath product $C_2 \wr C_d$). Let H be a CM-subgroup of G and $S \in \mathrm{CM}(G, H, \rho)$. It is generally known that

$$2 \log_2 |H\backslash G| \leq |r(S)\backslash G| \leq 2^{\frac{1}{2}|H\backslash G|}$$

and that there exists a CM-type S such that $|r(S)\backslash G| = 2^{\frac{1}{2}|H\backslash G|}$ holds (see [Ribet 1980, (3.2)]). In this example, we explicitly construct such CM-types when $\frac{1}{2}|H\backslash G|$ is odd.

Let d be an odd integer. We consider the wreath product $G = C_2 \wr C_d$, where C_d acts on d copies of C_2 by permutation. Hence the order of G is $2^d d$. The group G has a presentation

$$G = \langle c_1, \ldots, c_d, r \mid c_1^2 = \cdots c_d^2 = r^d = 1,\ r c_i r^{-1} = c_{i+1}\ (i = 1, \ldots, d) \rangle,$$

where the index i is understood modulo d. It is easy to show that $\rho = c_1 \cdots c_g$ is a central involution (in fact, a unique central involution). We consider the two subgroups of G

$$H = \langle c_2, c_3, \ldots, c_d \rangle, \quad K = \langle r \rangle.$$

They are obviously CM-subgroups of G with respect to ρ and we see $|H| = 2^{d-1}$ and $|K| = d$ and hence $|K \backslash G| = 2d = 2^{\frac{1}{2}|H \backslash G|}$ holds. We shall show $\mathscr{X}(H, K) = 2$.

We first show that H is a maximal CM-subgroup. Suppose that H' is a CM-subgroup such that $H' \gneqq H$. If $|H'|$ is a power of 2, then $|H'| = 2^d$ and H' is a 2-Sylow subgroup of G. On the other hand, we know that $C = \langle c_1, \ldots, c_g \rangle$ is a 2-Sylow group, which is normal in G. We thus conclude that $H' = C$ and $\rho \in H'$. This is a contradiction. Therefore there exists an odd prime p dividing $|H'|$ and, by Cauchy's theorem, there exists an element $x \in H'$ of order p. We can write $x = c r^k$ with $c \in H$ and an integer $d > k \geq 1$. We then have $x c_1 x^{-1} = c r^k c_1 r^{-k} c^{-1} = c c_{k+1} c^{-1} \in H$. Here we note that $c_{k+1} \neq c_1$. This implies $c_1 \in x^{-1} H x \subseteq H'$ and then $\rho \in H'$ and we again get a contradiction. Thus we have proved that H is a maximal CM-subgroup and therefore, we have

$$\mathscr{X}(H, K) = \mathscr{S}(H) \cap \mathscr{R}(K) = \mathscr{S}_{\geq}(H) \cap \mathscr{R}(K).$$

We use Proposition 3.3 to enumerate this.

To this end, we have to consider the groups N in $\mathscr{H}_{\geq}(K)$ and compute $|H \backslash G / N|$. We note that the cardinality $|HxK|$ of every double coset of G by H and N is divisible by both $|H|$ and $|K|$ and therefore we have $|HxK| = 2^{d-1} d$ or $2^d d$.

We begin with the case $N = K$. Let $c \in H$ and $r^k \in K$ and suppose that the order of $c r^k \in HK$ is 2. If we write $r^k c r^{-k} = c' \in C$, then we have $c r^k c r^k = c c' r^{2k} = 1$. This implies $2k \equiv 0 \pmod{d}$. We conclude that every element of order 2 in HK is contained in H. In particular, we obtain $\rho \notin HK$ and the double coset decomposition $G = HK \sqcup H\rho K$.

Next we consider $\mathscr{H}_{\geq}(K) \ni N \gneqq K$. There exists $c \in N$ of order 2, which is a product of some of $c_2, \ldots, c_g$. Since N is a subgroup, N also contains $r^k c r^{-k}$ for $0 \leq k < d$. At least one of the elements $r^k c r^{-k}$ contains c_1 as a cycle factor. This implies $\rho \in HN$ and we have $G = HN$.

Now it follows from Proposition 3.3 that

$$\mathscr{X}(H, K) = \sum_{N \in \mathscr{H}_{\geq}(K)} \varepsilon(H, N) \mu(K, N) 2^{\frac{1}{2}|H \backslash G / N|} = 2^{\frac{1}{2}|H \backslash G / K|} = 2$$

as desired.

The groups considered in Examples 6.3, 6.4, and 6.5 are all solvable groups. Thus the existence of CM-fields with Galois group isomorphic to these groups is guaranteed by Dodson's theorem [1986, Theorem 1.4]. Many explicit examples with small order are found in the database http://galoisdb.math.upb.de/. In particular, $C_2 \wr C_d$-extensions are constructed by starting from a totally real C_d-extension and using a construction explained in [Shimura 1970, 1.10].

References

[Aigner 2007] M. Aigner, *A course in enumeration*, Graduate Texts in Mathematics **238**, Springer, 2007. MR Zbl

[Dodson 1984] B. Dodson, “The structure of Galois groups of CM-fields”, *Trans. Amer. Math. Soc.* **283**:1 (1984), 1–32. MR Zbl

[Dodson 1986] B. Dodson, “Solvable and nonsolvable CM-fields”, *Amer. J. Math.* **108**:1 (1986), 75–93. MR Zbl

[Greenberg 1980] R. Greenberg, “On the Jacobian variety of some algebraic curves”, *Compositio Math.* **42**:3 (1980), 345–359. MR Zbl

[Oishi-Tomiyasu 2010] R. Oishi-Tomiyasu, “On some algebraic properties of CM-types of CM-fields and their reflexes”, *J. Number Theory* **130**:11 (2010), 2442–2466. MR Zbl

[Ribet 1980] K. A. Ribet, “Division fields of abelian varieties with complex multiplication”, pp. 75–94 in *Abelian functions and transcendental numbers* (Palaiseau, 1979), Mém. Soc. Math. France (N.S.) **2**, Soc. Math. France, Paris, 1980. MR Zbl

[Rota 1964] G.-C. Rota, “On the foundations of combinatorial theory, I: Theory of Möbius functions”, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **2** (1964), 340–368. MR Zbl

[Schappacher 1977] N. Schappacher, “Zur Existenz einfacher abelscher Varietäten mit komplexer Multiplikation”, *J. Reine Angew. Math.* **292** (1977), 186–190. MR Zbl

[Shimura 1970] G. Shimura, “On canonical models of arithmetic quotients of bounded symmetric domains”, *Ann. of Math.* (2) **91** (1970), 144–222. MR Zbl

Received 14 Mar 2019. Revised 17 May 2019.

MASANARI KIDA:

kida@rs.tus.ac.jp
Department of Mathematics, Faculty of Science Division I, Tokyo University of Science, Tokyo, Japan

Moscow Journal of Combinatorics and Number Theory
Vol. 8, No. 4, 2019
dx.doi.org/10.2140/moscow.2019.8.357

On polynomial-time solvable linear Diophantine problems

Iskander Aliev

We obtain a polynomial-time algorithm that, given input $(A, \boldsymbol{b})$, where $A = (B \mid N) \in \mathbb{Z}^{m\times n}$, $m < n$, with nonsingular $B \in \mathbb{Z}^{m\times m}$ and $\boldsymbol{b} \in \mathbb{Z}^m$, finds a nonnegative integer solution to the system $A\boldsymbol{x} = \boldsymbol{b}$ or determines that no such solution exists, provided that $\boldsymbol{b}$ is located sufficiently "deep" in the cone generated by the columns of B. This result improves on some of the previously known conditions that guarantee polynomial-time solvability of linear Diophantine problems.

1. Introduction and statement of results

Consider the linear Diophantine problem:

$$\begin{gathered}\text{Given } (A, \boldsymbol{b}), \text{ where } A \in \mathbb{Z}^{m\times n},\ m < n,\ \operatorname{rank}(A) = m \text{ and } \boldsymbol{b} \in \mathbb{Z}^m,\\ \text{find a nonnegative integer solution to the system } A\boldsymbol{x} = \boldsymbol{b}\\ \text{or determine that no such solution exists.}\end{gathered} \tag{1-1}$$

The problem (1-1) is referred to as the *multidimensional knapsack problem* and is NP-hard already for $m = 1$; see [Papadimitriou and Steiglitz 1982, Section 15.7].

Let $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in \mathbb{Z}^m$ be the columns of the matrix A and let

$$\mathcal{C}_A = \{\lambda_1\boldsymbol{v}_1 + \dots + \lambda_n\boldsymbol{v}_n : \lambda_1, \dots, \lambda_n \geq 0\}$$

be the cone generated by $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$. In this paper, we are interested in the problem of determining subsets $\mathcal{S} \subset \mathcal{C}_A$ such that (1-1) is solvable in polynomial time provided $\boldsymbol{b} \in \mathcal{S}$. We will use the general approach of [Gomory 1969], which was originally applied to study asymptotic integer programs, and combine it with results from discrete geometry.

We may assume, without loss of generality, that the matrix A is partitioned as

$$A = (B \mid N),$$

where $B \in \mathbb{Z}^{m\times m}$ is nonsingular and $N \in \mathbb{Z}^{m\times(n-m)}$. In what follows, we will denote by l_B and l_N the Euclidean lengths of the longest columns in the matrices B and N, respectively.

Let $\mathcal{C}_B \subset \mathcal{C}_A$ be the cone generated by the columns of the matrix B. The main result of this paper shows that (1-1) is solvable in polynomial time when the right-hand-side vector $\boldsymbol{b}$ is located deep enough in the cone $\mathcal{C}_B$.

Let $\mathcal{C}_B(t) \subset \mathcal{C}_B$ denote the affine cone of points in $\mathcal{C}_B$ at Euclidean distance $\geq t$ from the boundary of $\mathcal{C}_B$. We will denote by $\gcd(A)$ the greatest common divisor of all $m \times m$ subdeterminants of A.

MSC2010: primary 11D04, 90C10; secondary 11H06.

Keywords: multidimensional knapsack problem, polynomial-time algorithms, asymptotic integer programming, lattice points, Frobenius numbers.

Theorem 1.1. *There exists a polynomial-time algorithm which, given input* $(A, \boldsymbol{b})$, *where* $A = (B \mid N) \in \mathbb{Z}^{m\times n}$, *with nonsingular* $B \in \mathbb{Z}^{m\times m}$, *and*

$$\boldsymbol{b} \in \mathbb{Z}^m \cap \mathcal{C}_B\left(l_N\left(\frac{|\det(B)|}{\gcd(A)} - 1\right)\right), \tag{1-2}$$

finds a nonnegative integer solution to the system $A\boldsymbol{x} = \boldsymbol{b}$ *or determines that no such solution exists.*

We will now consider a special case where the matrix A satisfies the following conditions:

$$\begin{aligned} &\text{(i) } \gcd(A) = 1,\\ &\text{(ii) } \{\boldsymbol{x} \in \mathbb{R}^n_{\geq 0} : A\boldsymbol{x} = \boldsymbol{0}\} = \{\boldsymbol{0}\}. \end{aligned} \tag{1-3}$$

Notice that condition (i) in (1-3) guarantees that the system $A\boldsymbol{x} = \boldsymbol{b}$ has an integer solution for each $\boldsymbol{b} \in \mathbb{Z}^m$; see [Schrijver 1986, Corollary 4.1(c)]. The condition (ii) in (1-3) guarantees that the polyhedron $\{\boldsymbol{x} \in \mathbb{R}^n_{\geq 0} : A\boldsymbol{x} = \boldsymbol{b}\}$ is bounded.

When $m = 1$ in the setting (1-3), the problem (1-1) is linked to the well-known *Frobenius problem*; see [Ramírez Alfonsín 2005]. By condition (i) in (1-3), we have $\gcd(a_{11}, \dots, a_{1n}) = 1$ and by (ii) we may assume that the entries of A are positive. For such A the largest integer b such that (1-1) is infeasible is called the *Frobenius number* associated with A, denoted by $F(A)$. It is an interesting question to determine whether there exists a polynomial-time algorithm that solves (1-1) provided that

$$b > F(A);$$

see Conjecture 1.1 in [Aliev and Henk 2012].

The best known result in this direction is due to [Brimkov 1989]; see also [Aliev and Henk 2012; Brimkov 1988; Brimkov and Barneva 2001]. Specifically, set

$$f_1 = a_{11}, \qquad f_i = \gcd(a_{11}, \dots, a_{1i}), \quad i \in \{2, \dots, n\}. \tag{1-4}$$

A classical upper bound of [Brauer 1942] for the Frobenius numbers states that

$$F(A) \leq G(A) := a_{12}\frac{f_1}{f_2} + \cdots + a_{1n}\frac{f_{n-1}}{f_n} - \sum_{i=1}^{n} a_{1i}. \tag{1-5}$$

Brauer [1942] and, subsequently, Brauer and Seelbinder [1954] proved that the bound (1-5) is sharp and obtained a necessary and sufficient condition for the equality $F(A) = G(A)$. Brimkov [1989] gave a polynomial-time algorithm that solves (1-1) provided that

$$b > G(A). \tag{1-6}$$

We will show that an algorithm obtained in the proof of Theorem 1.1 matches the bound (1-6).

Corollary 1.2. *There exists a polynomial-time algorithm which, given input* (A, b), *where* $A \in \mathbb{Z}^{1\times n}_{>0}$ *satisfies* (1-3) *and* $b \in \mathbb{Z}$ *satisfies*

$$b > G(A),$$

computes a nonnegative integer solution to the equation $A\boldsymbol{x} = b$.

Recall that the *Minkowski sum* $X+Y$ of the sets $X, Y \subset \mathbb{R}^m$ consists of all points $\boldsymbol{x}+\boldsymbol{y}$ with $\boldsymbol{x} \in X$ and $\boldsymbol{y} \in Y$. For $m \geq 2$, Aliev and Henk [2012] considered the problem of estimating the minimal $t = t(A) \geq 0$ such that the problem (1-1) is solvable in polynomial time provided that A satisfies (1-3) and

$$\boldsymbol{b} \in \mathbb{Z}^m \cap (t\boldsymbol{v} + \mathcal{C}_A),$$

where $\boldsymbol{v} = \boldsymbol{v}_1 + \cdots + \boldsymbol{v}_n$ is the sum of columns of A.

Theorem 1.1 in [Aliev and Henk 2012] gives the bound

$$t \leq 2^{(n-m)/2-1} p(m,n)(\det(AA^T))^{1/2}, \tag{1-7}$$

where

$$p(m,n) = 2^{-1/2}(n-m)^{1/2} n^{1/2}.$$

Furthermore, Theorem 1.2 in [Aliev and Henk 2012] shows that the exponential factor $2^{(n-m)/2-1}$ in (1-7) is redundant for matrices with

$$\det(AA^T) > \frac{(n-m)2^{2(n-m-2)}\gamma_{n-m}^{n-m}}{n^2}. \tag{1-8}$$

Here γ_k is the k-dimensional Hermite constant, for which we refer to [Martinet 2003, Definition 2.2.5].

Let us now consider the case $m=2$. Condition (1-3)(ii) implies that the cone $\mathcal{C}_A$ is pointed. Thus we may assume without loss of generality that $A = (B \mid N)$ with $\mathcal{C}_B = \mathcal{C}_A$. The last result of this paper gives an estimate on the function $t(A)$ that is independent on the dimension n and allows a refinement of (1-7) when the ratio $l_B l_N / |\det(B)|$ is relatively small.

Corollary 1.3. *There exists a polynomial-time algorithm which, given input* $(A, \boldsymbol{b})$, *where* $A = (B \mid N) \in \mathbb{Z}^{2\times n}$, $B \in \mathbb{Z}^{2\times 2}$ *is nonsingular with* $\mathcal{C}_B = \mathcal{C}_A$, A *satisfies* (1-3) *and*

$$\boldsymbol{b} \in \mathbb{Z}^2 \cap \left(\frac{l_B l_N}{|\det(B)|}\Big(|\det(B)| - 1\Big)\boldsymbol{v} + \mathcal{C}_A\right), \tag{1-9}$$

computes a nonnegative integer solution to the system $A\boldsymbol{x} = \boldsymbol{b}$.

Noticing that $|\det(B)| \leq (\det(AA^T))^{1/2}$, condition (1-9) improves on (1-7) provided that $l_B l_N / |\det(B)| \leq 2^{(n-m)/2-1} p(m,n)$. For matrices A satisfying (1-8) an improvement occurs when $l_B l_N / |\det(B)| \leq p(m,n)$.

2. Tools from discrete geometry

For linearly independent $\boldsymbol{b}_1, \dots, \boldsymbol{b}_k$ in $\mathbb{R}^d$, the set $\Lambda = \left\{\sum_{i=1}^k \lambda_i \boldsymbol{b}_i : \lambda_i \in \mathbb{Z}\right\}$ is a k-dimensional *lattice* with *basis* $\boldsymbol{b}_1, \dots, \boldsymbol{b}_k$ and *determinant* $\det(\Lambda) = (\det(\boldsymbol{b}_i \cdot \boldsymbol{b}_j)_{1\leq i,j\leq k})^{1/2}$, where $\boldsymbol{b}_i \cdot \boldsymbol{b}_j$ is the standard inner product of the basis vectors $\boldsymbol{b}_i$ and $\boldsymbol{b}_j$. For a lattice $\Lambda \subset \mathbb{R}^d$ and $\boldsymbol{y} \in \mathbb{R}^d$, the set $\boldsymbol{y} + \Lambda$ is an *affine lattice* with determinant $\det(\Lambda)$.

Let Λ be a lattice in $\mathbb{R}^d$ with basis $\boldsymbol{b}_1, \dots, \boldsymbol{b}_d$ and let $\hat{\boldsymbol{b}}_i$ be the vectors obtained from the Gram–Schmidt orthogonalisation of $\boldsymbol{b}_1, \dots, \boldsymbol{b}_d$:

$$\hat{\boldsymbol{b}}_1 = \boldsymbol{b}_1, \qquad \hat{\boldsymbol{b}}_i = \boldsymbol{b}_i - \sum_{j=1}^{i-1} \mu_{i,j} \hat{\boldsymbol{b}}_j, \qquad j \in \{2, \dots, d\}, \tag{2-1}$$

where $\mu_{i,j} = (\boldsymbol{b}_i \cdot \hat{\boldsymbol{b}}_j)/|\hat{\boldsymbol{b}}_j|^2$.

We will associate with the basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d$ of Λ the box

$$\widehat{\mathcal{B}}(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d) = [0, \hat{\boldsymbol{b}}_1) \times [0, \hat{\boldsymbol{b}}_2) \times \cdots \times [0, \hat{\boldsymbol{b}}_d).$$

Lemma 2.1. *There exists a polynomial-time algorithm that, given a basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d$ of a d-dimensional lattice $\Lambda \subset \mathbb{Q}^d$ and a point $\boldsymbol{x}$ in $\mathbb{Q}^d$, finds a point $\boldsymbol{y} \in \Lambda$ such that $\boldsymbol{x} \in \boldsymbol{y} + \widehat{\mathcal{B}}(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)$.*

A proof of Lemma 2.1 is implicitly contained, for instance, in the description of the classical nearest-plane procedure of [Babai 1986]. For completeness, we include a proof that follows along an argument of the proof of Theorem 5.3.26 in [Grötschel et al. 1988].

Proof. Let $\boldsymbol{x}$ be any point of $\mathbb{Q}^d$. We need to find a point $\boldsymbol{y} \in \Lambda$ such that

$$\boldsymbol{x} - \boldsymbol{y} = \sum_{i=1}^{d} \lambda_i \hat{\boldsymbol{b}}_i, \quad \lambda_i \in [0, 1), \ i \in \{1, \ldots, d\}. \tag{2-2}$$

This can be achieved using the following procedure. First, we find the rational numbers λ_i^0, $i \in \{1, \ldots, d\}$, such that

$$\boldsymbol{x} = \sum_{i=1}^{d} \lambda_i^0 \hat{\boldsymbol{b}}_i.$$

This can be done in polynomial time by Theorem 3.3 in [Schrijver 1986]. Then we subtract $\lfloor \lambda_d^0 \rfloor \boldsymbol{b}_d$ to get a representation

$$\boldsymbol{x} - \lfloor \lambda_d^0 \rfloor \boldsymbol{b}_d = \sum_{i=1}^{d} \lambda_i^1 \hat{\boldsymbol{b}}_i,$$

where $\lambda_d^1 \in [0, 1)$. Next subtract $\lfloor \lambda_{d-1}^1 \rfloor \boldsymbol{b}_{d-1}$ and so on until we obtain the representation (2-2). □

Let now Λ be a d-dimensional sublattice of $\mathbb{Z}^d$. By Theorem I(A) and Corollary 1 in Chapter I of [Cassels 1959], there exists a unique basis $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_d$ of the sublattice Λ of the form

$$\begin{aligned} \boldsymbol{g}_1 &= v_{11}\boldsymbol{e}_1, \\ \boldsymbol{g}_2 &= v_{21}\boldsymbol{e}_1 + v_{22}\boldsymbol{e}_2, \\ &\ \vdots \\ \boldsymbol{g}_d &= v_{d1}\boldsymbol{e}_1 + \cdots + v_{dd}\boldsymbol{e}_d, \end{aligned} \tag{2-3}$$

where $\boldsymbol{e}_i$ are the standard basis vectors of $\mathbb{Z}^d$ and the coefficients v_{ij} satisfy the conditions $v_{ij} \in \mathbb{Z}$, $v_{ii} > 0$ for $i \in \{1, \ldots, d\}$ and $0 \le v_{ij} < v_{jj}$ for $i, j \in \{1, \ldots, d\}$, $i > j$.

Lemma 2.2. *There exists a polynomial-time algorithm that, given a basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d$ of a lattice $\Lambda \subset \mathbb{Z}^d$, finds the basis of Λ of the form (2-3).*

Proof. Let $V = (v_{ij}) \in \mathbb{Z}^{d \times d}$ be the matrix formed by the coefficients v_{ij} in (2-3) with $v_{ij} = 0$ for $j > i$. Observe that after a straightforward renumbering of the rows and columns of V we obtain a matrix in the row-style Hermite normal form. Now it is sufficient to notice that the Hermite normal form can be computed in polynomial time using an algorithm of [Kannan and Bachem 1979]. □

The Gram–Schmidt orthogonalisation (2-1) of the basis (2-3) of Λ has the form $\hat{\boldsymbol{g}}_1 = v_{11}\boldsymbol{e}_1, \dots, \hat{\boldsymbol{g}}_d = v_{dd}\boldsymbol{e}_d$. Therefore, noticing that the basis (2-3) is unique, we can associate with Λ the box

$$\mathcal{B}(\Lambda) = \widehat{\mathcal{B}}(\boldsymbol{g}_1, \dots, \boldsymbol{g}_d) = [0, v_{11}) \times [0, v_{22}) \times \cdots \times [0, v_{dd}).$$

Lemma 2.3. *For any $\boldsymbol{w} = (w_1, \dots, w_d)^T \in \mathcal{B}(\Lambda) \cap \mathbb{Z}^d$ we have*

$$\prod_{i=1}^{d}(1 + w_i) \leq \det(\Lambda).$$

Proof. It is sufficient to notice that by (2-3) $\det(\Lambda) = v_{11} \cdots v_{dd}$. □

3. Proof of Theorem 1.1

Given $A \in \mathbb{Z}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{Z}^m$, we will denote by $\Gamma(A, \boldsymbol{b})$ the set of integer points in the affine subspace

$$\mathcal{S}(A, \boldsymbol{b}) = \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} = \boldsymbol{b}\},$$

that is

$$\Gamma(A, \boldsymbol{b}) = \mathcal{S}(A, \boldsymbol{b}) \cap \mathbb{Z}^n.$$

The set $\Gamma(A, \boldsymbol{b})$ is either empty or is an affine lattice of the form $\Gamma(A, \boldsymbol{b}) = \boldsymbol{r} + \Gamma(A)$, where $\boldsymbol{r}$ is any integer vector with $A\boldsymbol{r} = \boldsymbol{b}$ and $\Gamma(A) = \Gamma(A, \boldsymbol{0})$ is the lattice formed by all integer points in the kernel of the matrix A. We will call the system $A\boldsymbol{x} = \boldsymbol{b}$ *integer feasible* if it has integer solutions or, equivalently, $\Gamma(A, \boldsymbol{b}) \neq \varnothing$. Otherwise the system is called *integer infeasible*.

Let π denote the projection map from $\mathbb{R}^n$ to $\mathbb{R}^{n-m}$ that forgets the first m coordinates. Recall that Theorem 1.1 applies to $A = (B \mid N)$, where B is nonsingular. It follows that the restricted map $\pi|_{\mathcal{S}(A,\boldsymbol{b})} : \mathcal{S}(A, \boldsymbol{b}) \to \mathbb{R}^{n-m}$ is bijective. Specifically, for any $\boldsymbol{w} \in \mathbb{R}^{n-m}$ we have

$$\pi|^{-1}_{\mathcal{S}(A,\boldsymbol{b})}(\boldsymbol{w}) = \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{w} \end{pmatrix}, \quad \text{with } \boldsymbol{u} = B^{-1}(\boldsymbol{b} - N\boldsymbol{w}).$$

For technical reasons, it is convenient to consider the projected set $\Lambda(A, \boldsymbol{b}) = \pi(\Gamma(A, \boldsymbol{b}))$ and the projected lattice $\Lambda(A) = \pi(\Gamma(A))$. Since the map $\pi|_{\mathcal{S}(A,\boldsymbol{0})}$ is bijective, we obtain the following lemma.

Lemma 3.1. *Let $\boldsymbol{g}_1, \dots, \boldsymbol{g}_{n-m}$ be a basis of $\Gamma(A)$. The vectors $\boldsymbol{b}_1 = \pi(\boldsymbol{g}_1), \dots, \boldsymbol{b}_{n-m} = \pi(\boldsymbol{g}_{n-m})$ form a basis of the lattice $\Lambda(A)$.*

Using notation of Lemma 3.1, let $G \in \mathbb{Z}^{n\times(n-m)}$ be the matrix with columns $\boldsymbol{g}_1, \dots, \boldsymbol{g}_{n-m}$. We will denote by F the $(n-m) \times (n-m)$-submatrix of G consisting of the last $n - m$ rows; hence, the columns of F are $\boldsymbol{b}_1, \dots, \boldsymbol{b}_{n-m}$. Then $\det(\Lambda(A)) = |\det(F)|$. The rows of the matrix A span the m-dimensional rational subspace of $\mathbb{R}^n$ orthogonal to the $(n-m)$-dimensional rational subspace spanned by the columns of G. Therefore, by Lemma 5G and Corollary 5I in [Schmidt 1991], we have $|\det(F)| = |\det(B)|/\gcd(A)$ and, consequently,

$$\det(\Lambda(A)) = \frac{|\det(B)|}{\gcd(A)}. \tag{3-1}$$

Consider the following algorithm.

Algorithm 1. *Input*: $(A, \boldsymbol{b})$, where $A = (B \mid N) \in \mathbb{Z}^{m\times n}$, $m < n$, with nonsingular $B \in \mathbb{Z}^{m\times m}$ and $\boldsymbol{b} \in \mathbb{Z}^m$.

Output: Solution $\boldsymbol{x} \in \mathbb{Z}^n$ to an integer feasible system $A\boldsymbol{x} = \boldsymbol{b}$.

Step 0: If $\Gamma(A, \boldsymbol{b}) = \varnothing$ then the system $A\boldsymbol{x} = \boldsymbol{b}$ is integer infeasible. Stop.

Step 1: Compute a point $\boldsymbol{z}$ of the affine lattice $\Lambda(A, \boldsymbol{b})$.

Step 2: Find a point $\boldsymbol{y} \in \Lambda(A)$ such that $\boldsymbol{z} \in \boldsymbol{y} + \mathcal{B}(\Lambda(A))$.

Step 3: Set $\boldsymbol{w} = \boldsymbol{z} - \boldsymbol{y}$ and output the vector

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{w} \end{pmatrix}, \quad \text{with } \boldsymbol{u} = B^{-1}(\boldsymbol{b} - N\boldsymbol{w}). \tag{3-2}$$

Note that Algorithm 1 will be also used in the proof of Corollary 1.2, where the condition (1-2) is replaced by its refinement (1-6). For this reason, we do not require that the input of the algorithm satisfies (1-2) and, as a consequence, the algorithm outputs a certain integer, but not necessarily nonnegative, solution to an integer feasible system $A\boldsymbol{x} = \boldsymbol{b}$ or detects integer infeasibility.

To complete the proof of Theorem 1.1, it is sufficient to show that Algorithm 1 is polynomial-time and that this algorithm computes a nonnegative integer solution to any integer feasible system $A\boldsymbol{x} = \boldsymbol{b}$ that satisfies its input conditions together with (1-2).

Let us show that all steps of Algorithm 1 can be computed in polynomial time. By Corollaries 5.3(b,c) in [Schrijver 1986] we can compute in polynomial time integer vectors $\boldsymbol{r}, \boldsymbol{g}_1, \dots, \boldsymbol{g}_{n-m}$ such that

$$\Gamma(A, \boldsymbol{b}) = \boldsymbol{r} + \sum_{i=1}^{n-m} \lambda_i \boldsymbol{g}_i, \quad \lambda_i \in \mathbb{Z},\ i \in \{1, \dots, n-m\}, \tag{3-3}$$

or determine that $\Gamma(A, \boldsymbol{b})$ is empty. This settles Steps 0 and 1. Further, the vectors $\boldsymbol{g}_1, \dots, \boldsymbol{g}_{n-m}$ in (3-3) form a basis of the lattice $\Gamma(A)$. In Step 2 we first find the projected vectors $\boldsymbol{b}_1 = \pi(\boldsymbol{g}_1), \dots, \boldsymbol{b}_{n-m} = \pi(\boldsymbol{g}_{n-m})$ that form a basis of the lattice $\Lambda(A)$ by Lemma 3.1. Then the point $\boldsymbol{y}$ can be computed in polynomial time using Lemmas 2.2 and 2.1. Finally, the lifted point $\boldsymbol{x}$ in Step 3 is computed in polynomial time by a straightforward calculation (3-2).

We will now show that Algorithm 1 computes a nonnegative integer solution to any integer feasible system $A\boldsymbol{x} = \boldsymbol{b}$ with $(A, \boldsymbol{b})$ satisfying its input conditions together with (1-2). By Step 0, we may assume that $\Gamma(A, \boldsymbol{b}) \neq \varnothing$ and hence at Step 1 we can find a point $\boldsymbol{z} \in \Lambda(A, \boldsymbol{b})$. At Step 2 we can find a point $\boldsymbol{y} \in \Lambda(A)$ with $\boldsymbol{z} \in \boldsymbol{y} + \mathcal{B}(\Lambda(A))$ by Lemma 2.1. Hence, the point $\boldsymbol{w} = \boldsymbol{z} - \boldsymbol{y}$ at Step 3 is a nonnegative point of the affine lattice $\Lambda(A, \boldsymbol{b})$. Further, since $\boldsymbol{w} \in \Lambda(A, \boldsymbol{b})$ and $\pi|_{\mathcal{S}(A,\boldsymbol{b})}$ is bijective, the point $\boldsymbol{x} = \pi|_{\mathcal{S}(A,\boldsymbol{b})}^{-1}(\boldsymbol{w})$ is integer. Summarising, we have

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{w} \end{pmatrix} \in \mathcal{S}(A, \boldsymbol{b}) \cap \mathbb{Z}^n \quad \text{and} \quad \pi(\boldsymbol{x}) = \boldsymbol{w} \geq \boldsymbol{0}. \tag{3-4}$$

It is now sufficient to show that $\boldsymbol{u} \geq \boldsymbol{0}$.

Observe that, by construction, $\boldsymbol{w} \in \mathcal{B}(\Lambda(A))$. Hence, Lemma 2.3, applied to $\boldsymbol{w}$ and $\Lambda = \Lambda(A)$, implies

$$\prod_{i=1}^{n-m} (1 + w_i) \leq \det(\Lambda(A)). \tag{3-5}$$

Expanding the product in (3-5) gives

$$\sum_{i=1}^{n-m} w_i \le \det(\Lambda(A)) - 1.$$

Hence, denoting by $\|\cdot\|_2$ the Euclidean norm, we obtain the inequality

$$\|N\boldsymbol{w}\|_2 \le l_N \sum_{i=1}^{n-m} w_i \le l_N(\det(\Lambda(A)) - 1). \tag{3-6}$$

By (3-1), $\boldsymbol{b} \in \mathcal{C}_B(l_N(\det(\Lambda(A)) - 1))$ and by (3-6), $\boldsymbol{b} - N\boldsymbol{w} \in \mathcal{C}_B$. The cone $\mathcal{C}_B$ can be written as

$$\mathcal{C}_B = \{\boldsymbol{y} \in \mathbb{R}^m : B^{-1}\boldsymbol{y} \ge \boldsymbol{0}\}$$

and therefore

$$\boldsymbol{u} = B^{-1}(\boldsymbol{b} - N\boldsymbol{w}) \ge \boldsymbol{0}.$$

□

4. Proof of Corollary 1.2

Let $A = (a_{11}, \dots, a_{1n}) \in \mathbb{Z}^{1\times n}$ satisfy (1-3). Then the lattice $\Lambda(A)$ can be written in the form

$$\Lambda(A) = \{\boldsymbol{x} \in \mathbb{Z}^{n-1} : a_{12}x_1 + \cdots + a_{1n}x_{n-1} \equiv 0 \pmod{a_{11}}\}.$$

Note also that $\det(\Lambda(A)) = a_{11}$ by (3-1).

The next lemma shows that the box $B(\Lambda(A))$ is entirely determined by the parameters f_i defined by (1-4).

Lemma 4.1. *The box $B = B(\Lambda(A))$ has the form*

$$B = \left[0, \frac{f_1}{f_2}\right) \times \left[0, \frac{f_2}{f_3}\right) \times \cdots \times \left[0, \frac{f_{n-1}}{f_n}\right).$$

Proof. By the definition of the box $B(\Lambda(A))$, it is sufficient to show that

$$v_{11} = \frac{f_1}{f_2}, \quad v_{22} = \frac{f_2}{f_3}, \quad \dots, \quad v_{n-1\,n-1} = \frac{f_{n-1}}{f_n}. \tag{4-1}$$

Let $\boldsymbol{g}_1, \dots, \boldsymbol{g}_{n-1}$ be the basis of the form (2-3) of the lattice $\Lambda(A)$. Let $\Lambda_i(A)$ denote the sublattice of $\Lambda(A)$ generated by the first i basis vectors $\boldsymbol{g}_1, \dots, \boldsymbol{g}_i$. We can write $\Lambda_i(A)$ in the form

$$\Lambda_i(A) = \left\{(x_1, \dots, x_i, 0, \dots, 0)^T \in \mathbb{Z}^{n-1} : \frac{a_{12}}{f_{i+1}}x_1 + \cdots + \frac{a_{1i+1}}{f_{i+1}}x_i \equiv 0 \left(\operatorname{mod} \frac{a_{11}}{f_{i+1}}\right)\right\}.$$

Hence, $\det(\Lambda_i(A)) = a_{11}/f_{i+1}$, $i \in \{1, \dots, n-1\}$. On the other hand, (2-3) implies

$$\det(\Lambda_i(A)) = v_{11}v_{22}\cdots v_{ii}, \quad i \in \{1, \dots, n-1\}.$$

Since $a_{11} = \det(\Lambda(A)) = v_{11}v_{22}\cdots v_{n-1\,n-1}$, we have

$$f_{i+1} = v_{i+1\,i+1}\cdots v_{n-1\,n-1} \quad \text{for } i \in \{1, \dots, n-2\}.$$

Noticing that $f_1 = a_{11}$ and $f_n = 1$, we obtain (4-1). □

Suppose that $b > G(A)$. Condition (1-3)(i) implies that the equation $A\boldsymbol{x} = b$ has integer solutions. Therefore, it is sufficient to show that the vector $\boldsymbol{x}$ computed by Algorithm 1 is nonnegative. When $m = 1$, (3-2) sets $\boldsymbol{x} = (u, w_1, \dots, w_{n-1})^T$ with

$$u = \frac{b - a_{12}w_1 - \cdots - a_{1n}w_{n-1}}{a_{11}}. \tag{4-2}$$

Further, (3-4) implies that $\boldsymbol{w} = (w_1, \dots, w_{n-1})^T \in \Lambda(A, b)$ is nonnegative and $u \in \mathbb{Z}$.

To see that $u \geq 0$, we observe first that the points of the affine lattice $\Lambda(A, b)$ are split into layers of the form

$$a_{12}x_1 + \cdots + a_{1n}x_{n-1} = b + ka_{11}, \quad k \in \mathbb{Z}. \tag{4-3}$$

Suppose, to derive a contradiction, that $u < 0$. Then, by (4-2),

$$a_{12}w_1 + \cdots + a_{1n}w_{n-1} > b. \tag{4-4}$$

On the other hand, by construction, $\boldsymbol{w} \in B(\Lambda(A))$ and hence, using Lemma 4.1 and noticing (1-5),

$$a_{12}w_1 + \cdots + a_{1n}w_{n-1} \leq G(A) + a_{11} < b + a_{11}. \tag{4-5}$$

Due to (4-3), the bounds (4-4) and (4-5) imply $\boldsymbol{w} \notin \Lambda(A, b)$. The obtained contradiction shows that $u \geq 0$.

5. Proof of Corollary 1.3

We will show that a nonnegative integer solution to the system $A\boldsymbol{x} = \boldsymbol{b}$ can be computed using Algorithm 1 from the proof of Theorem 1.1. By condition (1-3)(i), the system $A\boldsymbol{x} = \boldsymbol{b}$ is integer feasible. Following the proof of Theorem 1.1, it is sufficient to show that any $\boldsymbol{b}$ that satisfies (1-9) must satisfy (1-2).

Let h denote the distance from the vector $\boldsymbol{v}$ to the boundary of $\mathcal{C}_B$. Observe that we can write $\boldsymbol{v} = \boldsymbol{v}_1 + \boldsymbol{v}_2 + \boldsymbol{p}$, where $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ are the columns of B and $\boldsymbol{p} \in \mathcal{C}_B$. Therefore, we have

$$h \geq \frac{|\det(B)|}{l_B}$$

and, consequently, the points of the affine cone

$$\frac{l_B l_N}{|\det(B)|}(|\det(B)| - 1)\boldsymbol{v} + \mathcal{C}_A$$

are at the distance $\geq l_N(|\det(B)| - 1)$ to the boundary of $\mathcal{C}_B$.

Acknowledgement

The author is grateful to Valentin Brimkov, Martin Henk and Timm Oertel for valuable comments and suggestions.

References

[Aliev and Henk 2012] I. Aliev and M. Henk, "LLL-reduction for integer knapsacks", *J. Comb. Optim.* **24**:4 (2012), 613–626. MR Zbl

[Babai 1986] L. Babai, "On Lovász' lattice reduction and the nearest lattice point problem", *Combinatorica* **6**:1 (1986), 1–13. MR Zbl

[Brauer 1942] A. Brauer, "On a problem of partitions", *Amer. J. Math.* **64** (1942), 299–312. MR Zbl

[Brauer and Seelbinder 1954] A. Brauer and B. M. Seelbinder, "On a problem of partitions, II", *Amer. J. Math.* **76** (1954), 343–346. MR Zbl

[Brimkov 1988] V. E. Brimkov, "A polynomial algorithm for solving a large subclass of linear Diophantine equations in nonnegative integers", *C. R. Acad. Bulgare Sci.* **41**:11 (1988), 33–35. MR Zbl

[Brimkov 1989] V. E. Brimkov, "Effective algorithms for solving a broad class of linear Diophantine equations in nonnegative integers", pp. 241–246 in *Mathematics and education in mathematics* (Albena, Bulgaria, 1989), edited by G. Gerov, Bulgar. Akad. Nauk, Sofia, 1989. In Bulgarian. MR

[Brimkov and Barneva 2001] V. E. Brimkov and R. P. Barneva, "Gradient elements of the knapsack polytope", *Calcolo* **38**:1 (2001), 49–66. MR Zbl

[Cassels 1959] J. W. S. Cassels, *An introduction to the geometry of numbers*, Grundlehren der Math. Wissenschaften **99**, Springer, 1959. MR Zbl

[Gomory 1969] R. E. Gomory, "Some polyhedra related to combinatorial problems", *Linear Algebra and Appl.* **2** (1969), 451–558. MR Zbl

[Grötschel et al. 1988] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric algorithms and combinatorial optimization*, Algorithms Combinator. Study Res. Texts **2**, Springer, 1988. MR Zbl

[Kannan and Bachem 1979] R. Kannan and A. Bachem, "Polynomial algorithms for computing the Smith and Hermite normal forms of an integer matrix", *SIAM J. Comput.* **8**:4 (1979), 499–507. MR Zbl

[Martinet 2003] J. Martinet, *Perfect lattices in Euclidean spaces*, Grundlehren der Math. Wissenschaften **327**, Springer, 2003. MR Zbl

[Papadimitriou and Steiglitz 1982] C. H. Papadimitriou and K. Steiglitz, *Combinatorial optimization: algorithms and complexity*, Prentice-Hall, Englewood Cliffs, NJ, 1982. MR Zbl

[Ramírez Alfonsín 2005] J. L. Ramírez Alfonsín, *The Diophantine Frobenius problem*, Oxford Lect. Series Math. Appl. **30**, Oxford Univ. Press, 2005. MR Zbl

[Schmidt 1991] W. M. Schmidt, *Diophantine approximations and Diophantine equations*, Lecture Notes in Math. **1467**, Springer, 1991. MR Zbl

[Schrijver 1986] A. Schrijver, *Theory of linear and integer programming*, Wiley, Chichester, 1986. MR Zbl

Received 15 Mar 2019. Revised 1 Jul 2019.

ISKANDER ALIEV:
alievi@cardiff.ac.uk
Mathematics Institute, Cardiff University, Cardiff, United Kingdom

Moscow Journal of Combinatorics and Number Theory
Vol. 8, No. 4, 2019
dx.doi.org/10.2140/moscow.2019.8.367

Discrete analogues of John's theorem

Sören Lennart Berg and Martin Henk

As a discrete counterpart to the classical theorem of Fritz John on the approximation of symmetric n-dimensional convex bodies K by ellipsoids, Tao and Vu introduced so called generalized arithmetic progressions $\mathrm{P}(A, \boldsymbol{b}) \subset \mathbb{Z}^n$ in order to cover (many of) the lattice points inside a convex body by a simple geometric structure. Among others, they proved that there exists a generalized arithmetic progressions $\mathrm{P}(A, \boldsymbol{b})$ such that $\mathrm{P}(A, \boldsymbol{b}) \subset K \cap \mathbb{Z}^n \subset \mathrm{P}(A, O(n)^{3n/2}\boldsymbol{b})$. Here we show that this bound can be lowered to $n^{O(\ln n)}$ and study some general properties of so called unimodular generalized arithmetic progressions.

1. Introduction

Let $\mathcal{K}^n_{(s)}$ be the set of all o-symmetric convex bodies in $\mathbb{R}^n$, i.e., $K \in \mathcal{K}^n_{(s)}$ is a compact convex set in $\mathbb{R}^n$ with nonempty interior and $K = -K$. By $B_n \in \mathcal{K}^n_{(s)}$ we denote the n-dimensional Euclidean unit ball, i.e., $B_n = \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{x}, \boldsymbol{x}\rangle \leq 1\}$, where $\langle \cdot, \cdot \rangle$ is the standard inner product.

For $K \in \mathcal{K}^n_{(s)}$, John's (ellipsoid) theorem states that there exists an ellipsoid $\mathcal{E} \in \mathcal{K}^n_{(s)}$ such that

$$\mathcal{E} \subseteq K \subseteq \sqrt{n}\,\mathcal{E}. \tag{1-1}$$

(See, e.g., [Artstein-Avidan et al. 2015, Theorem 2.1.3] and [Schneider 2014, Theorem 10.12.2].) It turns out that the volume maximal ellipsoid contained in K gives the desired approximation, and in the nonsymmetric (or general) case the factor $\sqrt{n}$ has to be replaced by n (after a suitable translation of K).

This theorem has numerous applications in convex geometry or in the local theory of Banach spaces (see the two works just cited for examples). It allows one to get a first quick estimate on the value $f(K)$ of any homogenous and monotone functional f on $\mathcal{K}^n_{(s)}$ by the value of the functional at ellipsoids. For instance, if vol denotes the n-dimensional volume, i.e., n-dimensional Lebesgue measure, than (1-1) implies that for $K \in \mathcal{K}^n_{(s)}$ there exists an ellipsoid $\mathcal{E}$ such that

$$\operatorname{vol} \mathcal{E} \leq \operatorname{vol} K \leq n^{n/2} \operatorname{vol} \mathcal{E}. \tag{1-2}$$

In particular, the volume of an ellipsoid can easily be evaluated as $\mathcal{E} = A\,B_n$ for some $A \in \mathrm{GL}(n, \mathbb{R})$, and thus $\operatorname{vol} \mathcal{E} = |\det A| \operatorname{vol} B_n$.

Tao and Vu [2006] started to study a discrete version of John's theorem, where the aim of the approximation is the set of lattice points in K, i.e., the set $K \cap \mathbb{Z}^n$. The approximation itself is carried out

This paper contains some material from Berg's PhD thesis.

MSC2010: primary 11H06, 52C07; secondary 52A40.

Keywords: John's theorem, arithmetic progressions, convex bodies, lattices.

not by lattice points in ellipsoids, which are hard to control or to compute, but by a so called symmetric generalized arithmetic progression (GAP for short)

$$\mathrm{P}(A, \boldsymbol{b}) = \{A\,\boldsymbol{z} : \boldsymbol{z} \in \mathbb{Z}^n,\ |z_i| \le b_i,\ 1 \le i \le n\},$$

where $A \in \mathbb{Z}^{n\times n}$, $\det A \neq 0$, and $\boldsymbol{b} \in \mathbb{R}^n$. Hence, $\mathrm{P}(A, \boldsymbol{b})$ consists of the lattice points of the lattice $A\mathbb{Z}^n$ in the parallelepiped $\sum_{i=1}^n \operatorname{conv}\{-b_i\boldsymbol{a}_i, b_i\boldsymbol{a}_i\}$, where $\boldsymbol{a}_i$ is the i-th column of A and conv denotes the convex hull.

The same authors proved an improvement of an earlier result of theirs, [Tao and Vu 2006, Lemma 3.36]:

Theorem [Tao and Vu 2008, Theorem 1.6]. *Let $K \in \mathcal{K}^n_{(s)}$. There exists a GAP $\mathrm{P}(A, \boldsymbol{b}) \subset K$ such that*

$$K \cap \mathbb{Z}^n \subset \mathrm{P}(A, \mathrm{O}(n)^{3n/2}\boldsymbol{b}), \tag{1-3a}$$

$$|K \cap \mathbb{Z}^n| < \mathrm{O}(n)^{7n/2}|\mathrm{P}(A, \boldsymbol{b})|. \tag{1-3b}$$

(If C is a finite set, $|C|$ denotes its cardinality.) Observe that $|\mathrm{P}(A, \boldsymbol{b})| = \prod_{i=1}^n (2\lfloor b_i\rfloor + 1)$ can be easily computed. Obviously, (1-3a) and (1-3b) may be regarded as discrete counterparts to (1-1) and (1-2).

A first qualitative version of such a theorem, without mentioning explicit constants, was given in [Bárány and Vershik 1992, Theorem 3]. Here we prove:

Theorem 1.1. *Let $K \in \mathcal{K}^n_{(s)}$.*

(i) *There exists a GAP $\mathrm{P}(A, \boldsymbol{b}) \subset K$ such that*

$$K \cap \mathbb{Z}^n \subset \mathrm{P}(A, n^{\mathrm{O}(\ln n)}\,\boldsymbol{b}). \tag{1-4}$$

(ii) *There exists a GAP $\mathrm{P}(A, \boldsymbol{b}) \subset K$ such that*

$$|K \cap \mathbb{Z}^n| < \mathrm{O}(n)^n|\mathrm{P}(A, \boldsymbol{b})|. \tag{1-5}$$

In comparison to the volume case (John's ellipsoid) a GAP contained in $K \in \mathcal{K}^n_{(s)}$ that is optimal for the cardinality bound (1-5), i.e., covering most of the lattice points in K, does not need to be optimal for the inclusion bound (1-4) as well. We will give an example of this in Proposition 2.1. In fact, also the two GAPs leading to the bounds in (1-4) and (1-5) are different (in general).

Regarding a GAP $\mathrm{P}(A, \boldsymbol{b})$ which is simultaneously good with respect to inclusion and cardinality we have the following slight improvement on the above theorem of Tao and Vu.

Theorem 1.2. *Let $K \in \mathcal{K}^n_{(s)}$. There exists a GAP $\mathrm{P}(A, \boldsymbol{b}) \subset K$ such that*

$$K \cap \mathbb{Z}^n \subset \mathrm{P}(A, \mathrm{O}(n)^{2n/\ln n}\boldsymbol{b}), \tag{1-6a}$$

$$|K \cap \mathbb{Z}^n| < \mathrm{O}(n)^{2n}|\mathrm{P}(A, \boldsymbol{b})|. \tag{1-6b}$$

An *unconditional* convex body $K \in \mathcal{K}^n_{(s)}$ is one that is symmetric with respect to all coordinate hyperplanes. For such K, the inclusion bound can be made linear:

Proposition 1.3. *Let $K \in \mathcal{K}^n_{(s)}$ be an unconditional convex body. There exists a GAP* $\mathrm{P}(A, \boldsymbol{b}) \subset K$ *with*

$$K \cap \mathbb{Z}^n \subseteq \mathrm{P}(A, n\,\boldsymbol{b}), \tag{1-7a}$$

$$|K \cap \mathbb{Z}^n| < O(n)^n |\mathrm{P}(A, \boldsymbol{b})|. \tag{1-7b}$$

As we will show in Proposition 3.4, the linear inclusion bound in Proposition 1.3 is essentially best possible, and it might be even true that the bound of order $n^{\mathrm{O}(\ln n)}$ in (1-4) can be replaced by a linear or polynomial bound in n. In general, it seems to be a hard problem to construct explicitly a best possible GAP for one of the bounds; in fact, even the proofs yielding the results in the theorems above are rather nonconstructive. For unconditional bodies, however, the GAP behind the bounds in Proposition 1.3 can easily be described; see the proof of Proposition 1.3 on page 376 and the subsequent discussion.

For some other recent results regarding discretization of well-known inequalities from convex geometry we refer to, e.g., [Alexander et al. 2017; Hernández Cifre et al. 2018; Ryabogin et al. 2017].

The paper is organized as follows. In Section 2 we introduce and collect some basic properties of GAPs approximating the lattice points in symmetric convex bodies. In turns out that GAPs where the columns of A form a lattice basis of $\mathbb{Z}^n$ are of particular interest and we study them in Section 3. Finally, Section 4 contains the proofs of the theorems and of the proposition above.

2. Preliminaries and GAPs

For the proof of Theorem 1.1 it is more convenient to introduce GAPs for general lattices $\Lambda \subset \mathbb{R}^n$, i.e., $\Lambda = B\,\mathbb{Z}^n$, $B \in \mathbb{R}^{n\times n}$ with $\det B \neq 0$. Let $\mathcal{L}^n$ be the set of all these lattices. Following [Tao and Vu 2008], and adapting their definition to our special geometric situation, we define a generalized symmetric arithmetic progression with respect to Λ, or GAP, as the set of lattice points in Λ given by

$$\mathrm{P}(A, \boldsymbol{b}) = \{A\boldsymbol{z} : -\boldsymbol{b} \leq \boldsymbol{z} \leq \boldsymbol{b}, \boldsymbol{z} \in \mathbb{Z}^n\},$$

where $A \in \mathbb{R}^{n\times n}$ is a matrix with columns $\boldsymbol{a}_i \in \Lambda$, $1 \leq i \leq n$, and $\boldsymbol{b} \in \mathbb{R}^n_{>0}$.

Actually, Tao and Vu defined GAPs more generally, namely, for general $n \times m$ matrices A. In our geometric setting, however, this would make the inclusion bound needless as A may consist of all (up to ±) lattice points in $K \in \mathcal{K}^n_{(s)}$. Then, letting $\boldsymbol{b} = (1-\varepsilon)\boldsymbol{1}$, where $\boldsymbol{1}$ is the appropriate all 1-vector and ε an arbitrary positive number less than 1, gives the trivial inclusions

$$\{\boldsymbol{0}\} = P(A, \boldsymbol{b}) \subset K \cap \mathbb{Z}^n \subset P(A, (1-\varepsilon)^{-1}\boldsymbol{b})$$

Tao and Vu were mainly interested in so called infinitely proper GAPs which here means $m = \operatorname{rank}(A)$, and so we restrict the definition to the case $A \in \mathbb{R}^{n\times n}$, $\det A \neq 0$.

The size or cardinality of a GAP $\mathrm{P}(A, \boldsymbol{b})$ is given

$$|\mathrm{P}(A, \boldsymbol{b})| = \prod_{i=1}^{n} (2\lfloor b_i \rfloor + 1),$$

where $\lfloor \cdot \rfloor$ denotes the floor function. In general, for a vector $\boldsymbol{b} \in \mathbb{R}^n$ we denote by $\lfloor \boldsymbol{b} \rfloor = (\lfloor b_1 \rfloor, \dots, \lfloor b_n \rfloor)^\intercal$ its integral part. The parallelepiped associated to $\mathrm{P}(A, \boldsymbol{b})$ is denoted by

$$\mathrm{P}_{\mathbb{R}}(A, \boldsymbol{b}) = \{A\boldsymbol{x} : -\boldsymbol{b} \leq \boldsymbol{x} \leq \boldsymbol{b},\ \boldsymbol{x} \in \mathbb{R}^n\} = \sum_{i=1}^{n} \operatorname{conv}\{-b_i\boldsymbol{a}_i, b_i\boldsymbol{a}_i\}.$$

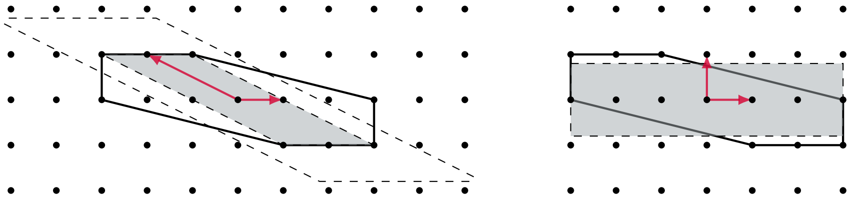
Observe that

$$\mathrm{P}_{\mathbb{R}}(A, \lfloor \boldsymbol{b} \rfloor) = \operatorname{conv} \mathrm{P}(A, \boldsymbol{b}). \tag{2-1}$$

Whenever we are interested in a GAP $\mathrm{P}(A, \boldsymbol{b})$ covering most of the lattice points in a convex body, i.e., a GAP which is optimal with respect to the cardinality bound, then it suffices to assume $\boldsymbol{b} \in \mathbb{N}^n$. However, for an optimal GAP with respect to the inclusion bound it might be essential to consider nonintegral vectors $\boldsymbol{b} \in \mathbb{R}^n_{>0}$. This is also reflected by the next example showing that those GAPs yielding an optimal cardinality bound can be different from those leading to an optimal inclusion bound.

Proposition 2.1. *Let $n \geq 2$. There exists a $K \in \mathcal{K}^n_{(s)}$ such that any GAP $\mathrm{P}(A, \boldsymbol{b}) \subset K$ covering most of the lattice points of K is not an optimal GAP with respect to inclusions, i.e., there exists another GAP $\mathrm{P}(\bar{A}, \bar{\boldsymbol{b}}) \subset K$ such that for any $t > 1$ with $K \cap \mathbb{Z}^n \subseteq \mathrm{P}(A, t\,\boldsymbol{b})$ there exits a $\bar{t} < t$ with $K \cap \mathbb{Z}^n \subseteq \mathrm{P}(\bar{A}, \bar{t}\,\bar{\boldsymbol{b}})$.*

Proof. We start with dimension 2, and let $K = \operatorname{conv}\{\pm(3, 0)^\intercal, \pm(-3, 1)^\intercal, \pm(-1, 1)^\intercal\}$, the hexagon in the figure.



We will argue that an optimal cardinality GAP $P(A, \boldsymbol{b}) \subseteq K$ contains 9 out of the 13 lattice points in K. To this end we may assume that the columns $\boldsymbol{a}_i$ of A belong to K, i.e., $\boldsymbol{a}_i \in K$ and $\boldsymbol{b} \geq \mathbf{1}$. Otherwise, we could only cover lattice points on a line which would be at most 7. Since for all $\boldsymbol{x} \in K$ we have $|x_2| \leq 1$, and since also the sum $\boldsymbol{a}_1 + \boldsymbol{a}_2$ has to belong to K, there is at most one column $\boldsymbol{a}_i$ of A having a nonzero last coordinate.

If there would be none, then again only the 7 points with last coordinate 0 could be covered.

Next assume that $\boldsymbol{a}_2$ is the vector having last coordinate nonzero and let $\boldsymbol{a}_1$ be the vector with last coordinate 0. The only possibility so that $\boldsymbol{a}_1 \pm \boldsymbol{a}_2$ belong to K is (up to sign) the one depicted in the left figure, i.e., $\boldsymbol{a}_1 = (1, 0)^\intercal$ and $\boldsymbol{a}_2 = (-2, 1)^\intercal$, and for any $\boldsymbol{b}$ with $1 \leq b_i < 2$, $i = 1, 2$, the GAP $\mathrm{P}(A, \boldsymbol{b})$ covers 9 out of the 13 lattice points of K. Hence, the GAPs covering the maximal amount of lattice points of K are given – up to ± and permutations of the columns of A – by $\mathrm{P}(A, \boldsymbol{b})$ for any $\boldsymbol{b}$ with $1 \leq b_i < 2$, $i = 1, 2$. Since $(3, 0)^\intercal \in K$, we observe that in order to cover all the points of $K \cap \mathbb{Z}^2$ by $\mathrm{P}(A, t\,\boldsymbol{b})$ we must have $t > \frac{3}{2}$.

On the other hand, if we take for the columns of $\bar{A}$ the vectors $(1, 0)^\intercal$ and $(0, 1)^\intercal$ and setting $\bar{\boldsymbol{b}} = (3, 1 - \varepsilon)^\intercal$ we get $|\mathrm{P}(\bar{A}, \bar{\boldsymbol{b}})| = 7$, but $K \cap \mathbb{Z}^2 \subset P(\bar{A}, (1 - \varepsilon)^{-1}\bar{\boldsymbol{b}})$ for any $\varepsilon \in (0, 1)$ (see the right half of the figure above).

This verifies the assertion in the plane. By building successively prisms over Q the example can be extended to all dimensions. □

3. Unimodular GAPs

Without loss of generality we consider here only the case $\Lambda = \mathbb{Z}^n$. The group of all unimodular matrices, i.e., integral $n \times n$-matrices of determinant ±1, is denoted by $\mathrm{GL}(n, \mathbb{Z})$; it consists of all lattice bases of

$\mathbb{Z}^n$. Apparently, if $K \cap \mathbb{Z}^n$ contains a lattice basis of $\mathbb{Z}^n$ and $K \cap \mathbb{Z}^n \subseteq \mathrm{P}(A, \boldsymbol{b})$ then $A \in \mathrm{GL}(n, \mathbb{Z})$. This basically shows that for the inclusion bound it suffices to consider GAPs $\mathrm{P}(U, \boldsymbol{b})$ with $U \in \mathrm{GL}(n, \mathbb{Z})$. We will call such a GAP an unimodular GAP.

Proposition 3.1. *Let $c = c(n) \in \mathbb{R}_{>0}$ be a constant depending on n. The following statements are equivalent.*

(i) *For every $K \in \mathcal{K}^n_{(s)}$ there exists a GAP $\mathrm{P}(A, \boldsymbol{b}) \subset K$ such that $K \cap \mathbb{Z}^n \subset \mathrm{P}(A, c\,\boldsymbol{b})$.*

(ii) *For every $K \in \mathcal{K}^n_{(s)}$ there exists an unimodular GAP $\mathrm{P}(U, \boldsymbol{b}) \subset K$ such that $K \cap \mathbb{Z}^n \subset \mathrm{P}(U, c\,\boldsymbol{b})$.*

Proof. Obviously, we only have to show that (i) implies (ii). To this end let $l \in \mathbb{N}$ such that $l\,K$ contains a basis of $\mathbb{Z}^n$. By assumption there exists a GAP $P(U, \boldsymbol{b}) \subset l K$ such that $l K \cap \mathbb{Z}^n \subseteq P(U, c\,\boldsymbol{b})$ and since $l K$ contains a basis of $\mathbb{Z}^n$ we have $U \in \mathrm{GL}(n, \mathbb{Z})$. Next we claim that

$$P(U, l^{-1}\boldsymbol{b}) \subseteq K \cap \mathbb{Z}^n \subseteq P(U, c\,l^{-1}\boldsymbol{b}). \tag{3-1}$$

Let $\boldsymbol{u} \in P(U, l^{-1}\boldsymbol{b})$. Then there exists a $\boldsymbol{z} \in \mathbb{Z}^n$ with $\boldsymbol{u} = U\boldsymbol{z}$ and $-l^{-1}\boldsymbol{b} \leq \boldsymbol{z} \leq l^{-1}\boldsymbol{b}$. Thus $l\boldsymbol{u} = U\,l\boldsymbol{z}$ and since $l\,\boldsymbol{z} \in \mathbb{Z}^n$ we get $l\boldsymbol{u} \in \mathrm{P}(U, \boldsymbol{b}) \subset l\,K$. Hence $\boldsymbol{u} \in K \cap \mathbb{Z}^n$ which shows the first inclusion in (3-1). For the second let $\boldsymbol{a} \in K \cap \mathbb{Z}^n$. Then $l\,\boldsymbol{a} \in l\,K \cap \mathbb{Z}^n \subseteq P(U, c\,\boldsymbol{b})$ and so there exists a $\boldsymbol{z} \in \mathbb{Z}^n$ with $-c\,\boldsymbol{b} \leq \boldsymbol{z} \leq c\,\boldsymbol{b}$ with $l\,\boldsymbol{a} = U\boldsymbol{z}$. Hence, $\boldsymbol{a} = U\,l^{-1}\boldsymbol{z}$ and since $U \in \mathrm{GL}(n, \mathbb{Z})$ we conclude $l^{-1}\boldsymbol{z} \in \mathbb{Z}^n$ which shows $\boldsymbol{a} \in P(U, c\,l^{-1}\boldsymbol{b})$. □

Next we want to point out a relation between GAPs and approximations of a convex body by an "unimodular" parallelepiped $\mathrm{P}_{\mathbb{R}}(U, \boldsymbol{u})$, $U \in \mathrm{GL}(n, \mathbb{Z})$. To this we first note that

Lemma 3.2. *Let $K \in \mathcal{K}^n_{(s)}$ containing n linearly independent points $\beta\boldsymbol{a}_i$ with $\beta \in \mathbb{R}_{>0}$ and $\boldsymbol{a}_i \in \mathbb{Z}^n$, $1 \leq i \leq n$. Then for any unimodular GAP $\mathrm{P}(U, \boldsymbol{u})$ with $K \subseteq \mathrm{P}_{\mathbb{R}}(U, \boldsymbol{u})$ we have $u_i \geq \beta$, $1 \leq i \leq n$.*

Proof. Let $\beta\boldsymbol{a}_i = U\boldsymbol{x}_i$ with $-\boldsymbol{u} \leq \boldsymbol{x}_i \leq \boldsymbol{u}$, $\boldsymbol{x}_i \in \mathbb{R}^n$. Since $U \in \mathrm{GL}(n, \mathbb{Z})$ we get $\boldsymbol{x}_i \in \beta\mathbb{Z}^n$, which shows that for each nonzero coordinate j, say, of $\boldsymbol{x}_i$ we have $u_j \geq \beta$. Since $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are linearly independent for each coordinate k we can find a vector $\boldsymbol{x}_l$ whose k-th coordinate is nonzero. □

Observe, for an unimodular GAP $\mathrm{P}(U, \boldsymbol{u})$ we have $\mathrm{P}(U, \boldsymbol{u}) = \mathrm{P}_{\mathbb{R}}(U, \boldsymbol{u}) \cap \mathbb{Z}^n$.

Proposition 3.3. *Let $c = c(n) \in \mathbb{R}_{>0}$ be a constant depending on n. The following statements are equivalent.*

(i) *For every $K \in \mathcal{K}^n_{(s)}$ there exists a GAP $\mathrm{P}(A, \boldsymbol{b}) \subset K$ such that $K \cap \mathbb{Z}^n \subseteq \mathrm{P}(A, c\,\boldsymbol{b})$.*

(ii) *For every $K \in \mathcal{K}^n_{(s)}$ there exists an unimodular GAP $\mathrm{P}(U, \boldsymbol{u}) \subset K$ such that*

$$\mathrm{P}_{\mathbb{R}}(U, \boldsymbol{u}) \subseteq K \subset \mathrm{P}_{\mathbb{R}}(U, c\,\boldsymbol{u}).$$

Proof. We start by showing that (i) implies (ii). Let $\varepsilon > 0$, and let $Q \subseteq K$ be a o-symmetric rational polytope with $K \subset (1+\varepsilon)Q$ (see, e.g., [Schneider 2014, Theorem 1.8.19]). Moreover, let $m \in \mathbb{N}$ be such that $m\,Q$ is an integral polytope (all its vertices are in $\mathbb{Z}^n$) and contains the scaled unit vectors $c(1+c/\varepsilon)\boldsymbol{e}_i$, $1 \leq i \leq n$. In view of Proposition 3.1 there exists an unimodular GAP $\mathrm{P}(U, \boldsymbol{u})$ such that

$$\mathrm{P}(U, \boldsymbol{u}) \subset m\,Q \cap \mathbb{Z}^n \subseteq \mathrm{P}(U, c\,\boldsymbol{u}).$$

The polytopes $\mathrm{P}_{\mathbb{R}}(U, \lfloor \boldsymbol{u} \rfloor)$ and mQ are integral and so we get

$$\begin{aligned} \mathrm{P}_{\mathbb{R}}(U, \lfloor \boldsymbol{u} \rfloor) &= \operatorname{conv}(\mathrm{P}_{\mathbb{R}}(U, \lfloor \boldsymbol{u} \rfloor) \cap \mathbb{Z}^n) = \operatorname{conv} \mathrm{P}(U, \lfloor \boldsymbol{u} \rfloor) \\ &\subseteq \operatorname{conv} \mathrm{P}(U, \boldsymbol{u}) \subseteq \operatorname{conv}(mQ \cap \mathbb{Z}^n) = mQ \subseteq mK. \end{aligned} \tag{3-2}$$

Since mQ is integral we have $mQ \subseteq \mathrm{P}_{\mathbb{R}}(U, c\,\boldsymbol{u})$ and due to Lemma 3.2 we know for the entries of $\boldsymbol{u}$ that $u_i \geq 1 + c/\varepsilon$, $1 \leq i \leq n$, which implies that

$$\frac{u_i}{\lfloor u_i \rfloor} \leq \frac{u_i}{u_i - 1} \leq 1 + \frac{\varepsilon}{c},$$

and thus $c\,\boldsymbol{u} \leq (c+\varepsilon)\lfloor \boldsymbol{u} \rfloor$. Hence,

$$\begin{aligned} mQ &= \operatorname{conv}(mQ \cap \mathbb{Z}^n) \subseteq \operatorname{conv} \mathrm{P}(U, c\,\boldsymbol{u}) \\ &\subseteq \mathrm{P}_{\mathbb{R}}(U, c\,\boldsymbol{u}) \subseteq \mathrm{P}_{\mathbb{R}}(U, (c+\varepsilon)\, \lfloor \boldsymbol{u} \rfloor), \end{aligned}$$

and with (3-2)

$$\mathrm{P}_{\mathbb{R}}(U, m^{-1}\lfloor \boldsymbol{u} \rfloor) \subseteq K \subseteq \mathrm{P}_{\mathbb{R}}(U, (1+\varepsilon)\,(c+\varepsilon)\, m^{-1}\lfloor \boldsymbol{u} \rfloor). \tag{3-3}$$

Observe, that actually $m = m_\varepsilon$, $U = U_\varepsilon$ as well as $\boldsymbol{u} = \boldsymbol{u}_\varepsilon$ depend on the chosen ε. Now, since K is bounded and all entries of U are integral, the first inclusion above shows that the sequence $m_\varepsilon^{-1}\lfloor \boldsymbol{u}_\varepsilon \rfloor$, $\varepsilon > 0$, has to be bounded. Therefore, we may assume that it converges to $\bar{\boldsymbol{u}}$ as ε approaches 0. Next assume that a sequence of a (fixed) column vector of the unimodular matrices U_ε is unbounded. Since $\operatorname{vol} \mathrm{P}_{\mathbb{R}}(U_\varepsilon, \mathbf{1}) = 2^n$ and since $m_\varepsilon^{-1}\lfloor \boldsymbol{u}_\varepsilon \rfloor$ is bounded this shows that the inradius of $\mathrm{P}_{\mathbb{R}}(U_\varepsilon, (1+\varepsilon)\,(c+\varepsilon)\, m_\varepsilon^{-1}\lfloor \boldsymbol{u}_\varepsilon \rfloor)$ must converge to 0 as ε tends to 0. This contradicts the second inclusion above and hence, also U_ε converges to an unimodular matrix $\overline{U}$. So we have shown

$$\mathrm{P}_{\mathbb{R}}(\overline{U}, \bar{\boldsymbol{u}}) \subseteq K \subseteq \mathrm{P}_{\mathbb{R}}(\overline{U}, c\,\bar{\boldsymbol{u}}).$$

For the reverse implication we assume that there exists an unimodular GAP $\mathrm{P}(U, \boldsymbol{u})$ fulfiling (ii). Then

$$\mathrm{P}_{\mathbb{R}}(U, \boldsymbol{u}) \cap \mathbb{Z}^n \subseteq K \cap \mathbb{Z}^n \subseteq \mathrm{P}_{\mathbb{R}}(U, c\,\boldsymbol{u}) \cap \mathbb{Z}^n,$$

and by the unimodularity of U we have $\mathrm{P}_{\mathbb{R}}(U, \boldsymbol{u}) \cap \mathbb{Z}^n = \mathrm{P}(U, \boldsymbol{u})$ as well as $\mathrm{P}_{\mathbb{R}}(U, c\,\boldsymbol{u}) \cap \mathbb{Z}^n = \mathrm{P}(U, c\,\boldsymbol{u})$. □

We close this section with lower bounds on the factors in (1-4) and (1-5) of Theorem 1.1.

Proposition 3.4.

(i) *Let $\tau = \tau(n) \in \mathbb{R}_{>0}$ be a constant depending on n such that for every $K \in \mathcal{K}^n_{(s)}$ there exists a GAP $\mathrm{P}(A, \boldsymbol{b}) \subset K$ such that $K \cap \mathbb{Z}^n \subseteq \mathrm{P}(A, \tau\,\boldsymbol{b})$. Then $\tau \geq n!^{1/n} > \frac{1}{\mathrm{e}} n$.*

(ii) *Let $\nu = \nu(n) \in \mathbb{R}_{>0}$ be a constant depending on n such that for every $K \in \mathcal{K}^n_{(s)}$ there exists a GAP $\mathrm{P}(A, \boldsymbol{b}) \subset K$ such that $|K \cap \mathbb{Z}^n| \leq \nu\,|\mathrm{P}(A, \boldsymbol{b})|$. Then $\nu \geq (2^n + 1)/3$.*

Proof. For (i) we consider for an integer $m \in \mathbb{N}$ the cross-polytope $mC_n^\star = \{\boldsymbol{x} \in \mathbb{R}^n : |x_1| + \cdots + |x_n| \leq m\}$ and let $\mathrm{P}(U, \boldsymbol{u})$ be a GAP such that

$$\mathrm{P}(U, \boldsymbol{u}) \subseteq mC_n^\star \cap \mathbb{Z}^n \subseteq \mathrm{P}(U, \tau\,\boldsymbol{u}). \tag{3-4}$$

In view of Proposition 3.1, or since $mC_n^\star$ contains the unit vectors $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ we have $U \in \mathrm{GL}(n, \mathbb{Z})$. Moreover, since $m\boldsymbol{e}_i \in mC_n^\star$, $1 \le i \le n$, we get from the second inclusion in (3-4) and Lemma 3.2 that $m \le \tau\, u_i$, $1 \le i \le n$, and so

$$\mathrm{vol}\,(mC_n^\star) = m^n \frac{2^n}{n!} \le \tau^n \frac{2^n}{n!} \prod_{i=1}^{n} u_i. \tag{3-5}$$

On the other hand, the first inclusion in (3-4) implies

$$\mathrm{P}_{\mathbb{R}}(U, \lfloor \boldsymbol{u} \rfloor) = \mathrm{conv}\,\mathrm{P}(U, \boldsymbol{u}) \subseteq mC_n^\star,$$

and so

$$2^n \prod_{i=1}^{n} \lfloor u_i \rfloor = \mathrm{vol}\,\mathrm{P}_{\mathbb{R}}(U, \lfloor \boldsymbol{u} \rfloor) \le \mathrm{vol}\,(mC_n^\star).$$

Combined with (3-5) we obtain

$$\tau \ge n!^{1/n} \left(\prod_{i=1}^{n} \frac{\lfloor u_i \rfloor}{u_i} \right)^{1/n}.$$

This is true for any $m \in \mathbb{N}$, and since $u_i \to \infty$ for $m \to \infty$, we find $\tau \ge n!^{1/n} > n/\mathrm{e}$.

To prove (ii), let Q be the o-symmetric lattice polytope given by $Q = \mathrm{conv}\,(\pm([0, 1]^{n-1} \times \{1\}))$. Then it is easy to see that $Q \cap \mathbb{Z}^n = \pm(\{0, 1\}^{n-1} \times \{1\}) \cup \{\boldsymbol{0}\}$ and hence, Q does not contain $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}^n \setminus \{\boldsymbol{0}\}$, $\boldsymbol{x} \ne -\boldsymbol{y}$, and $\boldsymbol{x} + \boldsymbol{y} \in Q$. Thus for any GAP $\mathrm{P}(A, \boldsymbol{b}) \subset Q$ we have $|\mathrm{P}(A, \boldsymbol{b})| \le 3$ and so

$$2^n + 1 = |Q \cap \mathbb{Z}^n| \le \nu\, |\mathrm{P}(A, \boldsymbol{b})| \le 3\,\nu,$$

yielding the desired lower bound. □

4. Proofs of the theorems

For the proof of the inclusion bound (1-4) of Theorem 1.1 we follow essentially the proof of [Tao and Vu 2008], but we apply a different lattice reduction taking into account also the polar lattice. More precisely, for a lattice $\Lambda \in \mathcal{L}^n$ with basis $B = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$, i.e., $\Lambda = B\mathbb{Z}^n$, we denote by

$$\Lambda^\star = \{\boldsymbol{y} \in \mathbb{R}^n : \langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \mathbb{Z} \text{ for all } \boldsymbol{x} \in \Lambda\} = B^{-\intercal}\mathbb{Z}^n$$

its polar lattice. In particular, if $B^{-T} = (\boldsymbol{b}_1^\star, \dots, \boldsymbol{b}_n^\star)$, then

$$\langle \boldsymbol{b}_i^\star, \boldsymbol{b}_j \rangle = \delta_{i,j}, \tag{4-1}$$

where $\delta_{i,j}$ denotes the Kronecker-symbol. Now a basis B of a lattice Λ is called Seysen reduced if

$$S(B) = \sum_{i=1}^{n} \|\boldsymbol{b}_i\|^2 \|\boldsymbol{b}_i^\star\|^2$$

is minimal among all bases of Λ (cf. [Seysen 1993]). Here, $\|\cdot\|$ denotes the Euclidean norm.

Theorem 4.1 [Seysen 1993, Theorem 7]. *Let $\Lambda \in \mathcal{L}^n$. There exists a basis $B = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$ of Λ such that $S(B) \le n^{O(\ln n)}$. In particular, for $1 \le i \le n$,*

$$\|\boldsymbol{b}_i\| \, \|\boldsymbol{b}_i^\star\| \le n^{O(\ln n)}. \tag{4-2}$$

For an explicit bound we refer to [Maze 2010] and for more information on lattice reduction and geometry of numbers we refer to [Gruber and Lekkerkerker 1987; Cassels 1959]. For the sake of comprehensibility we split the proof of Theorem 1.1 into two parts, one covering the inclusion bound and one the cardinality bound.

Proof of Theorem 1.1(i). In view of John's theorem (1-1) we may apply a linear transformation T to K such that with $\tilde{K} = TK$

$$B_n \subseteq \tilde{K} \subseteq \sqrt{n} B_n. \tag{4-3}$$

With $\Lambda = T\mathbb{Z}^n$ the problem is now to find a GAP $\mathrm{P}(A, \boldsymbol{b})$ in Λ such that $\mathrm{P}(A, \boldsymbol{b}) \subset \tilde{K}$ and

$$\tilde{K} \cap \Lambda \subset \mathrm{P}(A, n^{O(\ln n)}\boldsymbol{b}).$$

Let $B = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$ be a Seysen reduced basis of Λ with associated basis $B^{-\intercal} = (\boldsymbol{b}_1^\star, \dots, \boldsymbol{b}_n^\star)$ of the polar lattice and let $\boldsymbol{u} \in \mathbb{R}^n$ be given by $u_i = (1/n)\|\boldsymbol{b}_i\|^{-1}$, $1 \le i \le n$.

First, for $\boldsymbol{x} \in \mathrm{P}_{\mathbb{R}}(B, \boldsymbol{u})$ we have $\boldsymbol{x} = \sum_{i=1}^n \lambda_i \boldsymbol{b}_i$ with $|\lambda_i| \le u_i$ and by the triangle inequality we conclude $\|\boldsymbol{x}\| \le 1$. Hence, with (4-3) we certainly have $\mathrm{P}(B, \boldsymbol{u}) \subset \tilde{K}$. On the other hand, given $\boldsymbol{x} = \sum_{i=1}^n \beta_i \boldsymbol{b}_i \in \tilde{K}$ we get by Cramer's rule and (4-3)

$$|\beta_i| = \frac{|\det(\boldsymbol{x}, \boldsymbol{b}_1, \dots, \boldsymbol{b}_{i-1}, \boldsymbol{b}_{i+1}, \boldsymbol{b}_n)|}{|\det B|} \le \sqrt{n}\frac{\mathrm{vol}_{n-1}(\boldsymbol{b}_1, \dots, \boldsymbol{b}_{i-1}, \boldsymbol{b}_{i+1}, \boldsymbol{b}_n)}{\mathrm{vol}\,(\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)},$$

where $\mathrm{vol}_k(\boldsymbol{c}_1, \dots, \boldsymbol{c}_k)$ denotes the k-dimensional volume of the parallelepiped $\left\{\sum_{i=1}^k \mu_i \boldsymbol{c}_i : 0 \le \mu_i \le 1\right\}$. By (4-1) we find that

$$\mathrm{vol}\,(\boldsymbol{b}_1, \dots, \boldsymbol{b}_n) = \mathrm{vol}_{n-1}(\boldsymbol{b}_1, \dots, \boldsymbol{b}_{i-1}, \boldsymbol{b}_{i+1}, \boldsymbol{b}_n)\frac{\langle \boldsymbol{b}_i^\star, \boldsymbol{b}_i\rangle}{\|\boldsymbol{b}_i^\star\|} = \mathrm{vol}_{n-1}(\boldsymbol{b}_1, \dots, \boldsymbol{b}_{i-1}, \boldsymbol{b}_{i+1}, \boldsymbol{b}_n)\frac{1}{\|\boldsymbol{b}_i^\star\|},$$

and thus for $1 \le i \le n$

$$|\beta_i| \le \sqrt{n}\|\boldsymbol{b}_i^\star\|. \tag{4-4}$$

Together with the definition of u_i and Seysen's bound (4-2) we conclude that $|\beta_i| \le n^{3/2} n^{O(\ln n)} u_i$, for $1 \le i \le n$. Hence,

$$\tilde{K} \cap \Lambda \subseteq \mathrm{P}_{\mathbb{R}}(B, n^{O(\ln n)}\boldsymbol{u}) \cap \Lambda = \mathrm{P}(B, n^{O(\ln n)}\boldsymbol{u}),$$

since B is a basis of Λ. □

Remark 4.2. The optimal upper bound in Theorem 4.1 for a Seysen reduced basis is not known, but any improvement on this bound would immediately yield an improvement of (1-4).

For the cardinality bound (1-5) of Theorem 1.1 we need another tool from geometry of numbers: Minkowski's successive minima $\lambda_i(K, \Lambda)$, which for $K \in \mathcal{K}^n_{(s)}$, $\Lambda \in \mathcal{L}^n$ and $1 \le i \le n$ are defined by

$$\lambda_i(K, \Lambda) = \min\{\lambda > 0 : \dim(\lambda K \cap \Lambda) \ge i\}.$$

In words, $\lambda_i(K,\Lambda)$ is the smallest dilation factor λ such that $\lambda\, K$ contains i linearly independent lattice points of Λ. Minkowski's fundamental second theorem on successive minima (e.g., [Gruber and Lekkerkerker 1987, §9, Theorem 1]) states that

$$\operatorname{vol} K \le \det\Lambda \prod_{i=1}^{n} \frac{2}{\lambda_i(K,\Lambda)}, \tag{4-5}$$

and here we need a discrete version of it. In [Henk 2002] it was shown that

$$|K\cap\Lambda| \le 2^{n-1}\prod_{i=1}^{n}\left\lfloor \frac{2}{\lambda_i(K,\Lambda)}+1\right\rfloor, \tag{4-6}$$

and for an improvement on the constant 2^{n-1} and related results we refer to [Malikiosis 2010; Malikiosis 2012]. It is conjectured in [Betke et al. 1993] that (4-6) holds without any additional factor in front of the product which would, in particular, imply Minkowski's volume bound.

Proof of Theorem 1.1(ii). Let $\boldsymbol{a}_i\in\mathbb{Z}^n$, $1\le i\le n$, be linearly independent lattice vectors corresponding to the successive minima $\lambda_i=\lambda_i(K,\mathbb{Z}^n)$, i.e., $\boldsymbol{a}_i\in\lambda_i\, K$, $1\le i\le n$. Since $\lambda_i^{-1}\boldsymbol{a}_i\in K$ it follows

$$\left\{\sum_{i=1}^{n}\mu_i\frac{1}{n\lambda_i}\boldsymbol{a}_i : -1\le\mu_i\le 1\right\}\subset \operatorname{conv}\{\pm\lambda_i^{-1}\boldsymbol{a}_i : 1\le i\le n\}\subseteq K.$$

Thus, denoting by A the matrix with columns $\boldsymbol{a}_i$ and letting $\boldsymbol{b}$ be the vector with entries $b_i=(n\lambda_i)^{-1}$ we have $\mathrm{P}(A,\boldsymbol{b})\subset K$ and

$$|P(A,\boldsymbol{b})| = \prod_{i=1}^{n}\left(2\left\lfloor\frac{1}{n\lambda_i}\right\rfloor+1\right).$$

Now it is not hard to see that

$$2\left\lfloor\frac{1}{n\lambda_i}\right\rfloor+1 \ge \frac{1}{3}\frac{1}{n}\left\lfloor\frac{2}{\lambda_i}+1\right\rfloor, \tag{4-7}$$

and with (4-6) we get

$$|P(A,\boldsymbol{b})| \ge \left(\frac{1}{3n}\right)^{n}\left(\frac{1}{2}\right)^{n-1}2^{n-1}\prod_{i=1}^{n}\left\lfloor\frac{2}{\lambda_i}+1\right\rfloor > (6n)^{-n}|K\cap\mathbb{Z}^n|.$$

This shows (1-5). □

Remark 4.3. The columns of the matrix A of the GAP in the proof of the cardinality bound of Theorem 1.1 do not in general build a basis of $\mathbb{Z}^n$; hence this GAP cannot be used in order to obtain an inclusion bound.

Now the proof of Theorem 1.2 is a kind of combination of the two proofs leading to (1-4) and (1-5). Instead of a Seysen reduced basis we exploit properties of a so called Hermite–Korkin–Zolotarev (HKZ) reduced basis $\boldsymbol{b}_1,\dots,\boldsymbol{b}_n$ of the lattice Λ. For such a basis it was shown by Mahler (see, e.g., [Lagarias et al. 1990, Theorem 2.1]) that for $1\le i\le n$

$$\|\boldsymbol{b}_i\| \le \frac{\sqrt{i+3}}{2}\lambda_i(B_n,\Lambda). \tag{4-8}$$

Håstad and Lagarias [1990] pointed out that for a HKZ-basis one has

$$\|b_i\|\,\|b_i^\star\| \le \left(\tfrac{3}{2}\right)^n < n^{\frac{1}{2}n/\ln n}. \tag{4-9}$$

This bound is worse than the one given in (4-2), but the advantage of a HKZ reduced basis is its close relation to the successive minima (4-8).

Proof of Theorem 1.2. First we may assume that $\lambda_n(K,\mathbb{Z}^n)\le 1$, i.e., that K contains n linearly independent lattice points. Otherwise, all lattice points of K lying in a hyperplane H and it would be sufficient to prove the theorem with respect to the $n-1$-dimensional convex body $K\cap H$ and lattice $H\cap\mathbb{Z}^n$.

Now we proceed completely analogously to the proof of Theorem 1.1(i); we just replace the Seysen reduced basis by a HKZ-reduced basis $B=(\boldsymbol{b}_1,\dots,\boldsymbol{b}_n)$, and the GAP is given by $\mathrm{P}(B,\boldsymbol{u})$ with $u_i=(1/n)\|\boldsymbol{b}_i\|^{-1}$, $1\le i\le n$. Replacing (4-2) by (4-9) in (4-4) leads then to

$$\mathrm{P}(B,\boldsymbol{u})\subseteq\tilde{K}\cap\Lambda\subseteq\mathrm{P}(B,n^{O(n/\ln n)}\boldsymbol{u}),$$

where $\tilde{K}$ was a linear image of K such that

$$B_n\subseteq\tilde{K}\subseteq\sqrt{n}B_n. \tag{4-10}$$

It remains to prove the cardinality bound for the GAP $\mathrm{P}(B,\boldsymbol{u})$ and $\tilde{K}$. Regarding the size of $\mathrm{P}(B,\boldsymbol{u})$ we have

$$|\mathrm{P}(B,\boldsymbol{u})|=\prod_{i=1}^n\left(2\left\lfloor\frac{1}{n\|\boldsymbol{b}_i\|}\right\rfloor+1\right)\ge n^{-n}\prod_{i=1}^n\frac{1}{\|\boldsymbol{b}_i\|}. \tag{4-11}$$

On the other hand, for an upper bound on $\tilde{K}\cap\Lambda$ we use (4-6) and since $\lambda_n(K,\Lambda)\le 1$ we get

$$|K\cap\Lambda|\le 2^{n-1}\prod_{i=1}^n\left(\frac{2}{\lambda_i(K,\Lambda)}+1\right)\le 6^n\prod_{i=1}^n\frac{1}{\lambda_i(K,\Lambda)}.$$

In view of (4-10) and (4-8) we obtain

$$|K\cap\Lambda|\le 6^n\prod_{i=1}^n\frac{1}{\lambda_i(\sqrt{n}B_n,\Lambda)}=(6\sqrt{n})^n\prod_{i=1}^n\frac{1}{\lambda_i(B_n,\Lambda)}\le(6n)^n\prod_{i=1}^n\frac{1}{\|\boldsymbol{b}_i\|}.$$

Combined with (4-11) we get $|K\cap\Lambda|\le O(n)^{2n}|\mathrm{P}(B,\boldsymbol{u})|$. □

Next we consider unconditional bodies $K\in\mathcal{K}^n_{(s)}$, i.e., bodies which are symmetric with respect to all coordinate hyperplanes. As stated in Proposition 1.3, in this special case the inclusion bound can be made linear in the dimension. In view of Proposition 3.4 this is also the optimal order within this class of bodies as the given example used for the lower bound in Proposition 3.4 is unconditional.

Proof of Proposition 1.3. For $i=1,\dots,n$ let u_i be the maximal entry of the i-th coordinate of a point of K. Then $u_i>0$ and

$$K\cap\mathbb{Z}^n\subseteq\mathrm{P}(I_n,\boldsymbol{u}) \tag{4-12}$$

with $\boldsymbol{u}=(u_1,\dots,u_n)^\intercal$ and I_n the $n\times n$-identity matrix. By the unconditionality of K we have $\pm u_i\,\boldsymbol{e}_i\in K$, $1\le i\le n$, and thus

$$\mathrm{P}_{\mathbb{R}}(I_n,n^{-1}\boldsymbol{u})\subset\operatorname{conv}\{\pm u_i\boldsymbol{e}_i:1\le i\le n\}\subseteq K.$$

Hence, $\mathrm{P}(I_n, n^{-1}\boldsymbol{u}) \subset K$. For the remaining cardinality bound observe that $(2\,u_i+1) < (2\lfloor u_i/n\rfloor)+1)\,3n$ and so (4-12) implies

$$|K\cap\mathbb{Z}^n| \le \prod_{i=1}^{n}(2\,\lfloor u_i\rfloor+1) < (3n)^n\prod_{i=1}^{n}(2\lfloor u_i/n\rfloor)+1) = (3n)^n|\mathrm{P}(I_n, n^{-1}\boldsymbol{u})|. \qquad \square$$

For instance, for $p\ge 1$ and a positive vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^\intercal \in \mathbb{R}^n_{>0}$ let

$$B_n^p(\boldsymbol{\alpha}) = \Big\{\boldsymbol{x}\in\mathbb{R}^n : \sum_{i=1}^{n}\alpha_i^{-p}|x_i|^p \le 1\Big\}$$

be the scaled l_p-ball in $\mathbb{R}^n$. Then, by the preceding argument we get

$$\mathrm{P}(I_n, n^{-1/p}\boldsymbol{\alpha}) \subset B_n^p(\boldsymbol{\alpha}) \subset \mathrm{P}(I_n, \boldsymbol{\alpha}).$$

Assuming $\alpha_1\ge\alpha_2\ge\cdots\ge\alpha_n$ we also have $\lambda_i(B_n^p(\boldsymbol{\alpha}), \mathbb{Z}^n) = \alpha_i^{-1}$, and so the GAP corresponds to the vectors $\alpha_i\boldsymbol{e}_i \in K$ attaining the successive minima (compare Remark 4.3).

Finally, we remark that for a symmetric planar convex body K there always exists vectors $\boldsymbol{a}_1, \boldsymbol{a}_2 \in \mathbb{Z}^2$ such that $\boldsymbol{a}_i \in \lambda_i(K, \mathbb{Z}^2)\,K$, $i = 1, 2$, and $\boldsymbol{a}_1, \boldsymbol{a}_2$ build a basis of $\mathbb{Z}^2$. Setting $A = (\boldsymbol{a}_1, \boldsymbol{a}_2)$, it can be shown (see [Berg 2018, Theorem 4.21]) that there exists a GAP $\mathrm{P}(A, \boldsymbol{u}) \subset K$ satisfying

$$K\cap\mathbb{Z}^n \subseteq \mathrm{P}(A, 3\,\boldsymbol{u}).$$

It is not known, however, whether the dilation factor 3 is optimal.

References

[Alexander et al. 2017] M. Alexander, M. Henk, and A. Zvavitch, "A discrete version of Koldobsky's slicing inequality", *Israel J. Math.* **222**:1 (2017), 261–278. MR Zbl

[Artstein-Avidan et al. 2015] S. Artstein-Avidan, A. Giannopoulos, and V. D. Milman, *Asymptotic geometric analysis, I*, Math. Surv. Monogr. **202**, Amer. Math. Soc., Providence, RI, 2015. MR Zbl

[Bárány and Vershik 1992] I. Bárány and A. M. Vershik, "On the number of convex lattice polytopes", *Geom. Funct. Anal.* **2**:4 (1992), 381–393. MR Zbl

[Berg 2018] S. L. Berg, *Lattice points in convex bodies: counting and approximating*, Ph.D. thesis, Technische Universität Berlin, 2018, Available at https://tinyurl.com/bergconv.

[Betke et al. 1993] U. Betke, M. Henk, and J. M. Wills, "Successive-minima-type inequalities", *Discrete Comput. Geom.* **9**:2 (1993), 165–175. MR Zbl

[Cassels 1959] J. W. S. Cassels, *An introduction to the geometry of numbers*, Grundlehren der Math. Wissenschaften **99**, Springer, 1959. MR Zbl

[Gruber and Lekkerkerker 1987] P. M. Gruber and C. G. Lekkerkerker, *Geometry of numbers*, 2nd ed., North-Holland Math. Library **37**, North-Holland, Amsterdam, 1987. MR Zbl

[Håstad and Lagarias 1990] J. Håstad and J. C. Lagarias, "Simultaneously good bases of a lattice and its reciprocal lattice", *Math. Ann.* **287**:1 (1990), 163–174. MR Zbl

[Henk 2002] M. Henk, "Successive minima and lattice points", *Rend. Circ. Mat. Palermo* (2) *Suppl.* **70**:1 (2002), 377–384. MR Zbl

[Hernández Cifre et al. 2018] M. A. Hernández Cifre, D. Iglesias, and J. Yepes Nicolás, "On a discrete Brunn–Minkowski type inequality", *SIAM J. Discrete Math.* **32**:3 (2018), 1840–1856. MR Zbl

[Lagarias et al. 1990] J. C. Lagarias, H. W. Lenstra, Jr., and C.-P. Schnorr, "Korkin–Zolotarev bases and successive minima of a lattice and its reciprocal lattice", *Combinatorica* **10**:4 (1990), 333–348. MR Zbl

[Malikiosis 2010] R. Malikiosis, "An optimization problem related to Minkowski's successive minima", *Discrete Comput. Geom.* **43**:4 (2010), 784–797. MR Zbl

[Malikiosis 2012] R.-D. Malikiosis, "A discrete analogue for Minkowski's second theorem on successive minima", *Adv. Geom.* **12**:2 (2012), 365–380. MR Zbl

[Maze 2010] G. Maze, "Some inequalities related to the Seysen measure of a lattice", *Linear Algebra Appl.* **433**:8-10 (2010), 1659–1665. MR Zbl

[Ryabogin et al. 2017] D. Ryabogin, V. Yaskin, and N. Zhang, "Unique determination of convex lattice sets", *Discrete Comput. Geom.* **57**:3 (2017), 582–589. MR Zbl

[Schneider 2014] R. Schneider, *Convex bodies: the Brunn–Minkowski theory*, 2nd expanded ed., Encycl. Math. Appl. **151**, Cambridge Univ. Press, 2014. MR Zbl

[Seysen 1993] M. Seysen, "Simultaneous reduction of a lattice basis and its reciprocal basis", *Combinatorica* **13**:3 (1993), 363–376. MR Zbl

[Tao and Vu 2006] T. Tao and V. Vu, *Additive combinatorics*, Cambridge Studies in Adv. Math. **105**, Cambridge Univ. Press, 2006. MR Zbl

[Tao and Vu 2008] T. Tao and V. Vu, "John-type theorems for generalized arithmetic progressions and iterated sumsets", *Adv. Math.* **219**:2 (2008), 428–449. MR Zbl

Received 14 Apr 2019. Revised 27 May 2019.

SÖREN LENNART BERG:

berg@math.tu-berlin.de
Institut für Mathematik, Technische Universität Berlin, Berlin, Germany

MARTIN HENK:

henk@math.tu-berlin.de
Institut für Mathematik, Technische Universität Berlin, Berlin, Germany

Moscow Journal of Combinatorics and Number Theory
Vol. 8, No. 4, 2019
dx.doi.org/10.2140/moscow.2019.8.379

On the domination number of a graph defined by containment

Peter Frankl

Let $n > k > 2$ be integers. Define a bipartite graph between all k-element and all 2-element subsets of an n-element set by drawing an edge if and only if the first one contains the second. The domination number of this graph is determined up to a factor of $1 + o(1)$. The short proof relies on some extremal results concerning hypergraphs.

1. Introduction

For a graph $\mathcal{G} = (V, \mathcal{E})$ a subset $D \subset V$ is called a *dominating set* if for every vertex $x \in V \setminus D$ there is an edge $E \in \mathcal{E}$ satisfying $x \in E$ and $E \cap D \neq \varnothing$. The *domination number* $\varrho(\mathcal{G})$ is the minimum of $|D|$ over all dominating sets.

To determine $\varrho(\mathcal{G})$ for a given graph is very difficult in general. In the present paper we address this problem for a bipartite graph defined via containments of sets.

For n and k positive integers, with $n > k$, we denote by $[n] = \{1, 2, \dots, n\}$ the standard n-element set and by $\binom{[n]}{k}$ the collection of all k-element subsets of $[n]$. For integers $n > k > \ell \geq 2$, we define the bipartite graph $\mathcal{B} = \mathcal{B}_n(k, \ell)$ on the vertex set $\binom{[n]}{k} \cup \binom{[n]}{\ell}$ by drawing an edge between $F \in \binom{[n]}{k}$ and $G \in \binom{[n]}{\ell}$ if and only if $G \subset F$.

The problem of determining or estimating $\varrho(\mathcal{B})$ was raised in [Badakhshian et al. 2019] by Badakhshian, Katona and Tuza. They determined $\varrho(\mathcal{B}_n(3, 2))$ up to a factor $1 + o(1)$, where $o(1) \to 0$ as $n \to \infty$.

In the present paper we extend their work to all $k \geq 3$.

Theorem 1.1.
$$\varrho\bigl(\mathcal{B}_n(k, 2)\bigr) = (1 + o(1))\binom{n}{2}\frac{k+3}{k^2-1}.$$

To prove the lower bound we use a result from [Erdős et al. 1986] extending the celebrated Ruzsa–Szemerédi theorem [1978]. To obtain the matching upper bound we apply a probabilistic construction based on a result of [Frankl and Rödl 1985]. To prove similar results for $\varrho\bigl(\mathcal{B}_n(k, \ell)\bigr)$ where $\ell \geq 3$ appears to be much harder (Section 4).

2. Proof of the lower bound

Let $k \geq 3$ be fixed and $\varepsilon > 0$ be arbitrarily small. Choose $\mathcal{G} \subset \binom{[n]}{2}$ and $\mathcal{F} \subset \binom{[n]}{k}$ such that $\mathcal{F} \cup \mathcal{G}$ is a dominating set for $\mathcal{B} = \mathcal{B}_n(k, 2)$. Our aim is to prove

$$|\mathcal{F}| + |\mathcal{G}| > \binom{n}{2}\left(\frac{k+3}{k^2-1} - \varepsilon\right). \tag{1}$$

This research was done while the author was visiting Academia Sinica in Taipei.
Keywords: finite sets, graphs, hypergraphs, Turán's theorem.

Since $\frac{k+3}{k^2-1} \leq \frac{3}{4}$ for $k \geq 3$, we may assume that

$$|\mathcal{F}| \leq \frac{3}{4}\binom{n}{2}. \tag{2}$$

Proposition 2.1. $$|\mathcal{G}| > \frac{1-\varepsilon}{k-1}\binom{n}{2} \quad \textit{for all } n > n_0(k,\varepsilon).$$

Proof of the proposition. Let m be an integer (later qualified) and consider an m-element set $R \subset [n]$. If R contains no $F \in \mathcal{F}$, then the assumption on domination is equivalent to the fact that $\mathcal{G}_{|R} := \mathcal{G} \cap \binom{R}{2}$ has no independent set of k vertices. By Turán's theorem [1941] (or see [Bollobás 1978]), we have

$$\begin{aligned}\left|\mathcal{G} \cap \binom{R}{2}\right| > (k-1)\binom{m/(k-1)}{2} &= \frac{m(m-k+1)}{2(k-1)} \\ &> \binom{m}{2}\frac{1-\varepsilon/2}{k-1} \quad \text{for } m > 2k/\varepsilon. \end{aligned} \tag{3}$$

We now assume m is large enough that (3) is satisfied. Let us choose the set $P \in \binom{[n]}{m}$ uniformly at random.

Claim 2.2. *Let $n > m^3/\varepsilon$. Then the probability of $\binom{P}{k} \cap \mathcal{F} \neq \varnothing$ is smaller than $\varepsilon/2$.*

Proof. Since each $F \in \mathcal{F}$ is contained in $\binom{n-k}{m-k}$ subsets $R \in \binom{[n]}{m}$, (2) implies the upper bound $\frac{3}{4}\binom{n}{2}\binom{n-k}{m-k}$ on the number of R in question. Using $k \geq 3$ we obtain the upper bound

$$\frac{3}{4}\binom{n}{2}\binom{n-3}{m-3} = \binom{n}{m} \cdot \frac{m-2}{n-2}\binom{m}{2} \cdot \frac{3}{4} < \binom{n}{m}\frac{m^3}{2n} < \frac{\varepsilon}{2}\binom{n}{m}.$$

In view of the claim, for $n > m^3/\varepsilon$ a proportion of more than $(1-\varepsilon/2)$ of $R \in \binom{[n]}{m}$ satisfy (3). Now $(1-\varepsilon/2)^2 > 1-\varepsilon$ implies the inequality in Proposition 2.1, with $n_0(k,\varepsilon) > (2k/\varepsilon)^3/\varepsilon$. □

Let $\mathcal{H} = \binom{[n]}{2} \setminus \mathcal{G}$ be the graph of those edges $H \in \binom{[n]}{2}$ that are not in $\mathcal{G}$. Since $\mathcal{F} \cup \mathcal{G}$ is a dominating set for $\mathcal{B}$, for each $H \in \mathcal{H}$ there exists some $F \in \mathcal{F}$ with $H \subset F$. From this we infer

$$|\mathcal{F}| \geq \frac{|\mathcal{H}|}{\binom{k}{2}}. \tag{4}$$

Using (4) together with Proposition 2.1 one can show that

$$|\mathcal{F}| + |\mathcal{G}| \geq \frac{1-\varepsilon}{k-1}\binom{n}{2} + \frac{k-2+\varepsilon}{(k-1)} \frac{\binom{n}{2}}{\binom{k}{2}}$$

which is slightly weaker than (1). To prove (1), we would need (4) with $\binom{k}{2} - 1$ in the denominator.

Our strategy is relatively simple. We try and list (some of) the edges of $\mathcal{F}$: $F_1, F_2, \ldots, F_q$ such that $\binom{F_1}{2} \cap \mathcal{G} \neq \varnothing$, then $\binom{F_2}{2} \cap \left(\mathcal{G} \cup \binom{F_1}{2}\right) \neq \varnothing$, etc. That is, we choose sequentially F_i, $1 \leq i \leq q$, so that $\binom{F_i}{2} \cap \mathcal{G} \neq \varnothing$ or $|F_j \cap F_i| \geq 2$ for some $1 \leq j < i$. For each F_i let $\mathcal{E}(F_i)$ consist of those $E \in \mathcal{H}$ that $E \notin F_j$ for $1 \leq j < i$. From the construction it follows that

$$\left|\mathcal{E}(\mathcal{F}_i)\right| \leq \binom{k}{2} - 1 \quad \text{for all } 1 \leq i \leq q. \tag{5}$$

Should $\mathcal{F} = \{F_1, \ldots, F_q\}$ hold, (1) would follow. In the opposite case set $\mathcal{F}_0 = \{F_1, \ldots, F_q\}$ and $\mathcal{H}_0 = \left(\binom{F_1}{2} \cup \ldots \cup \binom{F_q}{2}\right) \setminus \mathcal{G}$.

Choosing q maximal, $\binom{F}{2} \cap \mathcal{G} = \varnothing$ and $|F \cap F_i| \le 1$ follow for $F \in \mathcal{F} \setminus \mathcal{F}_0$, $1 \le i \le q$.

We define $\mathcal{F}_1 = \{F_1, \ldots, F_{q_1}\}$ similarly. We choose $F_1 \in \mathcal{F} \setminus \mathcal{F}_0$ arbitrarily and once $F_1, \ldots, F_{s-1} \in \mathcal{F} \setminus \mathcal{F}_0$ are fixed, we choose an arbitrary $F_s \in \mathcal{F} \setminus \mathcal{F}_0$ from the rest, satisfying $|F_i \cap F_s| \ge 2$ for some $1 \le j < s$. Now let $\mathcal{F}_1$ be a maximal collection obtained in this way. This choice guarantees $|F \cap F'| \le 1$ for all $F \in \mathcal{F} \setminus (\mathcal{F}_0 \cup \mathcal{F}_1)$, $F' \in \mathcal{F}_1$.

Set $\mathcal{H}_1 = \bigcup\limits_{F \in \mathcal{F}_1} \binom{F}{2}$. Our procedure guarantees

$$|\mathcal{H}_1| \le 1 + |\mathcal{F}_1| \left(\binom{k}{2} - 1\right). \tag{6}$$

We iterate this procedure. Once $\mathcal{F}_1, \ldots, \mathcal{F}_p$ and thereby $\mathcal{H}_i = \bigcup\limits_{F \in \mathcal{F}_i} \binom{F}{2}$, $1 \le i \le p$ are chosen we have

$$|F \cap F'| \le 1 \quad \text{for all } F \in \mathcal{G} \cup \mathcal{F}_0 \cup \ldots \cup \mathcal{F}_p \text{ and } F' \in \mathcal{F} \setminus (\mathcal{F}_0 \cup \ldots \cup \mathcal{F}_p).$$

As long as there are sets remaining in $\mathcal{F}$ we can define $\mathcal{F}_{p+1}$ and $\mathcal{H}_{p+1}$ in the above way.

Eventually we obtain a partition,

$$\mathcal{F} = \mathcal{F}_0 \sqcup \ldots \sqcup \mathcal{F}_t$$

such that

$$\mathcal{H}_0 \sqcup \ldots \sqcup \mathcal{H}_t = \binom{[n]}{2} \setminus \mathcal{G}$$

(here we used that $\mathcal{G} \cup \mathcal{F}$ is a dominating set). Moreover (6) holds for 1 replaced by i:

$$|\mathcal{H}_i| \le 1 + |\mathcal{F}_i| \left(\binom{k}{2} - 1\right), \qquad 1 \le i \le t. \tag{7}$$

Since for $i = 0$ we do not need the extra 1, we infer

$$\binom{n}{2} - |\mathcal{G}| \le t + |\mathcal{F}| \left(\binom{k}{2} - 1\right),$$

or equivalently

$$|\mathcal{G}| + |\mathcal{F}| \ge \frac{\binom{n}{2}}{\binom{k}{2} - 1} + |\mathcal{G}| \frac{\binom{k}{2} - 2}{\binom{k}{2} - 1} - \frac{t}{\binom{k}{2} - 1}.$$

Substituting $|\mathcal{G}| > \frac{1-\varepsilon}{k-1}\binom{n}{2}$ we obtain

$$\begin{aligned} |\mathcal{G}| + |\mathcal{F}| &> \frac{\binom{n}{2}}{\binom{k}{2} - 1} \left(1 + \frac{\binom{k}{2} - 2}{k-1} - \frac{\varepsilon}{k-1}\right) - \frac{t}{\binom{k}{2} - 1} \\ &= \binom{n}{2} \left(\frac{k+3}{k^2 - 1} - \frac{2\varepsilon}{(k^2-1)(k-2)}\right) - \frac{t}{\binom{k}{2} - 1}. \end{aligned}$$

To conclude the proof of the lower bound it is clearly more than sufficient to show that $t = o\left(\binom{n}{2}\right)$. To achieve this we will need the following extension of a celebrated result from [Ruzsa and Szemerédi 1978]:

Theorem 2.3 (Erdős, Frankl, Rödl [Erdős et al. 1986]). *Suppose that $\mathcal{T} \subset \binom{[n]}{k}$ satisfies $|T \cap T'| \le 1$ for all distinct $T, T' \in \mathcal{T}$, moreover one cannot find a k-set $\{x_1, \dots, x_k\} \subset [n]$ and $\binom{k}{2}$ distinct members $T(i, j) \in \mathcal{T}$, $1 \le i < j \le k$, such that $\{x_i, x_j\} \subset T(i, j)$. Then*

$$|\mathcal{T}| = o\left(\binom{n}{2}\right). \tag{8}$$

To apply (8) we choose $F(i)$ as an arbitrary member of $\mathcal{F}_i$ for $1 \le i \le t$ and define

$$\mathcal{T} = \{F(i) : 1 \le i \le t\}.$$

The condition $|T \cap T'| \le 1$ is automatically satisfied. To prove the second condition we argue indirectly.

Suppose that we found $F = \{x_1, \dots, x_k\}$ and $\binom{k}{2}$ members $T(i, j) \in \mathcal{T}$ such that $\{x_i, x_j\} \subset T(i, j)$. Since $\mathcal{F} \cup \mathcal{G}$ is a dominating set for B, either $F \in \mathcal{F}$ or $G \subset F$ for some $G \in \mathcal{G}$. In the latter case $G = \{x_i, x_j\}$ for some $1 \le i < j \le k$. I.e., $G \subset T(i, j)$. But this is impossible since we put all such $T(i, j)$ into $\mathcal{F}_0$. Suppose next $F \in \mathcal{F}$. Assume by symmetry $T(1, 2) \in \mathcal{F}_1$, $T(1, 3) \in \mathcal{F}_2$. From $|T(1, \ell) \cap F| \ge 2$ we infer $F \in \mathcal{F}_{\ell-1}$ for $\ell = 2, 3$. This is impossible because of $\mathcal{F}_1 \cap \mathcal{F}_2 = \varnothing$, giving the desired contradiction. □

3. The proof of the upper bound

We give a probabilistic construction based on the following old result.

Let $r \ge 2$ be an integer and consider an r-uniform hypergraph $\mathcal{H} \subset \binom{X}{r}$, where $|X| = m$. For $x \in X$ let $d(x)$ be the degree of x in $\mathcal{H}$, that is, the number of $H \in \mathcal{H}$ containing x. The *double degree* $d(x, y)$ is defined analogously.

The *covering index* $b(\mathcal{H})$ is defined as the minimal number b such that there exist b edges in $\mathcal{H}$ whose union is equal to X. Obviously, $b(\mathcal{H}) \ge m/r$.

Theorem 3.1 [Frankl and Rödl 1985]. *Let β, ε be positive constants, $r \ge 2$ fixed. There exists $\delta = \delta(r, \beta, \varepsilon)$ such that, for every $\mathcal{H} \subset \binom{X}{r}$ satisfying*

(i) $\big|d(x) - |\mathcal{H}|r/m\big| < \delta|\mathcal{H}|/m$ ***or***

(ii) $d(x, y) < |\mathcal{H}|r/m^{1+\beta}$,

one has $b(\mathcal{H}) < (1 + \varepsilon)m/r$.

Now we are ready to explain the construction of a nearly optimal dominating set for $\mathcal{B}_n(k, 2)$, $k \ge 3$. (Badakhshian et al. [2019] use the same construction for the case $k = 3$.)

Let $n = p(k-1) + q$, $0 \le q < k - 1$ and let $[n] = X_1 \sqcup \ldots \sqcup X_{k-1}$ be a partition with $p \le |X_i| \le p + 1$. Let $\mathcal{G} := \bigcup_{1 \le i < k} \binom{X_i}{2}$ be the so-called *Turán graph*. By the pigeonhole principle, $\mathcal{G}$ dominates all k-sets in $\mathcal{B}_n(k, 2)$.

Set $r = \binom{k}{2} - 1$. We define an r-uniform hypergraph $\mathcal{H}$ on the partite set $\binom{[n]}{2}$ from $\mathcal{B}_n(k, 2)$. Note that for every k-set $F \subset [n]$ satisfying $F \cap X_i \neq \varnothing$ for $1 \le i < k$ there is exactly one $j = j(F)$ such that $|F \cap X_j| = 2$. With such an F we associate the r-set $H(F) = \binom{F}{2} \setminus \{F \cap X_j\}$. Let $\mathcal{H}$ be the r-graph formed by these $H(F)$. The actual vertex set of $\mathcal{H}$ is

$$X = \binom{[n]}{2} \setminus \left(\binom{X_1}{2} \cup \ldots \cup \binom{X_{k-1}}{2} \right);$$

that is, the number of vertices is $m \sim \frac{k-2}{k-1}\binom{n}{2}$.

If $|X_1| = \cdots = |X_{k-1}|$, then $\mathcal{H}$ is regular but even in the general case it is nearly regular. That is, (i) holds for $m > m(\delta)$.

Since $|\mathcal{H}| = (k-1+o(1))p^k/2$ and $|\mathcal{H}(x, y)| < p^{k-3}$, (ii) is satisfied with e.g. $\beta = \frac{1}{3}$ if $m > m_0(k, \beta)$.

Applying Theorem 3.1 we obtain a covering of X which is, say, formed by the edges $H(F_1), \ldots, H(F_b)$, $b < (1+\varepsilon)m/r$.

Let $\mathcal{F} = \{F_1, \ldots, F_b\}$ be the corresponding family in $\binom{[n]}{k}$. Then $\mathcal{G} \cup \mathcal{F}$ is a dominating set for $\mathcal{B}_n(k, 2)$. Substituting $m = (1+o(1))\frac{k-2}{k-1}\binom{n}{2}$, $r = \binom{k}{2} - 1$, we infer

$$|\mathcal{G} \cup \mathcal{F}| \le \binom{n}{2}\left(\frac{1}{k-1} + \frac{k-2}{k-1} \cdot \frac{1}{\binom{k}{2} - 1} + \varepsilon\right) = \binom{n}{2}\left(\frac{k+3}{k^2-1} + \varepsilon\right).$$

Since $\varepsilon > 0$ was arbitrary, this concludes the proof of the upper bound in Theorem 1.1. □

4. The general problem

Let us say a few words about $\varrho\big(\mathcal{B}_n(k, \ell)\big)$ in the case $\ell \ge 3$. One would imagine that to find a small dominating set imitating the strategy used for $\ell = 2$ should be the best. However, that means that first we choose $\mathcal{G} \subset \binom{[n]}{\ell}$ covering the whole of $\binom{[n]}{k}$, that is, for every $F \in \binom{[n]}{k}$ there exists $G \in \mathcal{G}$ with $G \subset F$.

The problem is that we do not know the minimal size, $|\mathcal{G}|$ for such families. It is the famous Turán's Problem (cf. [Turán 1961]) which is still open for all pairs (k, ℓ), $k > \ell \ge 3$.

At the same time there are some plausible conjectures. For example Turán [Turán 1961] conjectured that in the case $k = 5$, $\ell = 3$ and $n > n_0(k, \ell)$ the best construction is $\mathcal{G} = \binom{X}{3} \cup \binom{Y}{3}$ where $X \cup Y = [n]$ is a partition and $|X| = \lfloor \frac{n}{2} \rfloor$. Using this $\mathcal{G}$ one can use the approach of Section 3 and show that

$$\varrho\big(\mathcal{B}_n(5, 3)\big) \le (1+o(1))\left(\frac{1}{4} + \frac{3}{4}\frac{1}{\binom{5}{3} - 1}\right)\binom{n}{3} = \left(\frac{1}{3} + o(1)\right)\binom{n}{3}. \tag{9}$$

Using the results of [Frankl and Rödl 2002] one can prove the matching lower bound assuming that Turán's conjecture is true.

The situation is pretty much the same for other pairs (k, ℓ) whenever the conjectured optimal family for Turán's Problem is a "highly regular" ℓ-graph.

Let us close this paper with a conjecture.

Conjecture 4.1. $\varrho\big(\mathcal{B}_n(2\ell-1, \ell)\big) = (1+o(1))\left(\frac{1}{2^{\ell-1}} + \left(1 - \frac{1}{2^{\ell-1}}\right) \Big/ \left(\binom{2\ell-1}{\ell} - 1\right)\right)\binom{n}{3}.$

Acknowledgement

The author is indebted to Professor G. O. H. Katona for telling him about the problem and fruitful discussions.

References

[Badakhshian et al. 2019] L. Badakhshian, G. O. H. Katona, and Z. Tuza, "The domination number of the graph defined by two levels of the n-cube", *Discrete Appl. Math.* **266** (2019), 30–37. MR

[Bollobás 1978] B. Bollobás, *Extremal graph theory, with emphasis on probabilistic methods*, London Math. Soc. Monographs **11**, Academic Press, 1978. Reprinted Dover, Mineola (NY), 2004. MR

[Erdős et al. 1986] P. Erdős, P. Frankl, and V. Rödl, "The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent", *Graphs Combin.* **2**:2 (1986), 113–121. MR

[Frankl and Rödl 1985] P. Frankl and V. Rödl, "Near perfect coverings in graphs and hypergraphs", *European J. Combin.* **6**:4 (1985), 317–326. MR Zbl

[Frankl and Rödl 2002] P. Frankl and V. Rödl, "Extremal problems on set systems", *Random Structures Algorithms* **20**:2 (2002), 131–164. MR

[Ruzsa and Szemerédi 1978] I. Z. Ruzsa and E. Szemerédi, "Triple systems with no six points carrying three triangles", pp. 939–945 in *Combinatorics: Proceedings of the Fifth Hungarian Colloquium* (Keszthely, 1976), vol. 2, Colloq. Math. Soc. János Bolyai **18**, North-Holland, Amsterdam, 1978. MR Zbl

[Turán 1941] P. Turán, "On an extremal problem in graph theory", *Mat. Fiz. Lapok* **48** (1941), 436–452. In Hungarian. MR Zbl

[Turán 1961] P. Turán, "Research problems", *Publ. Math. Inst. Hungar. Acad. Sci.* **6** (1961), 417–423. MR

Received 29 Apr 2019. Revised 1 Aug 2019.

PETER FRANKL:
peter.frankl@gmail.com
Rényi Institute, Budapest, Hungary

Moscow Journal of Combinatorics and Number Theory
Vol. 8, No. 4, 2019
dx.doi.org/10.2140/moscow.2019.8.385

A new explicit formula for Bernoulli numbers involving the Euler number

Sumit Kumar Jha

We derive a new explicit formula for Bernoulli numbers in terms of the Stirling numbers of the second kind and the Euler numbers. As a corollary of our result, we obtain an explicit formula for the even Euler numbers in terms of the Stirling numbers of the second kind.

Definition 1. The *Bernoulli numbers* B_n can be defined by the generating function

$$\frac{t}{e^t-1}=\sum_{n\geq 0}\frac{B_n t^n}{n!},$$

where $|t|<2\pi$.

Definition 2. A *Stirling number of the second kind*, denoted by $S(n,m)$, is the number of ways of partitioning a set of n elements into m nonempty sets.

There are many known explicit formulas known for the Bernoulli numbers [Gould 1972; Jha 2019]. The following formulas express the Bernoulli numbers explicitly in terms of the Stirling numbers of the second kind:

$$B_r=\sum_{k=1}^{r}(-1)^k\cdot k!\,\frac{S(r,k)}{k+1},$$

$$(-1)^{r-1}B_r=\sum_{k=1}^{r}(-1)^k\frac{S(r,k)}{k+1}\cdot(k-1)!\,,$$

$$B_{r+1}=\frac{(-1)^r\cdot(r+1)\cdot 2^r}{2^{r+1}-1}\sum_{k=1}^{r}\frac{S(r,k)}{k+1}(-1)^k 2^{-2k}\frac{(2k-1)!}{(k-1)!}.$$

Definition 3. The *Euler numbers* are a sequence of integers, denoted by E_n, which can be defined by the Taylor series expansion

$$\frac{1}{\cosh t}=\frac{2}{e^t+e^{-t}}=\sum_{n=0}^{\infty}\frac{E_n}{n!}\cdot t^n,$$

where $\cosh t$ is the hyperbolic cosine.

We prove the following.

MSC2010: 11B68.
Keywords: Bernoulli numbers, Stirling numbers of the second kind, Euler numbers, polylogarithm function.

Theorem 4. *We have*

$$B_{r+1} = -\frac{r+1}{4(1+2^{-(r+1)}(1-2^{-r}))}\left(\sum_{k=1}^{r}(-1)^k \cdot \frac{S(r,k)}{k+1} \cdot \left(\tfrac{3}{4}\right)^{(k)} + 4^{-r}E_r\right), \tag{1}$$

where $S(r,k)$ denotes the Stirling numbers of the second kind, $x^{(n)} = (x)(x+1)\cdots(x+n-1)$ denotes the rising factorial, and E_r denotes the Euler number.

Proof. We begin with the result

$$\frac{\sin n\pi}{\pi}\int_0^\infty x^{n-1}\frac{\operatorname{Li}_s(-x)}{1+x}\,dx = \zeta(s) - \zeta(s, 1-n),$$

where $\operatorname{Li}_s(-x)$ denotes the polylogarithm function, $\zeta(s)$ is the Riemann zeta function, and $\zeta(s, 1-n)$ is the Hurwitz zeta function. The integral above is valid for all $s \in \mathbb{C}\setminus\{1\}$ and $0 < n < 1$. This integral can be obtained from formula 3.2.1.6 in [Brychkov et al. 2019].

Plugging $n = \frac{3}{4}$ and $s = -r$, a negative integer, into the integral above we get

$$\int_0^\infty x^{-1/4}\frac{\operatorname{Li}_{-r}(-x)}{1+x}\,dx = \sqrt{2}\pi\left(\frac{B_{r+1}\left(\frac{1}{4}\right) - B_{r+1}}{r+1}\right).$$

Now, we use the representation from [Landsburg 2009]

$$\operatorname{Li}_{-r}(-x) = \sum_{k=1}^{r} k!\, S(r,k)\left(\frac{1}{1+x}\right)^{k+1}(-x)^k,$$

which can be easily proved using induction on r.

As a result, we have

$$\begin{aligned}
\int_0^\infty x^{-1/4}\frac{\operatorname{Li}_{-r}(-x)}{1+x}\,dx &= \sum_{k=1}^{r}(-1)^k \cdot k!\, S(r,k)\int_0^\infty \frac{x^{k-1/4}}{(1+x)^{k+2}}\,dx \\
&= \sum_{k=1}^{r}(-1)^k \cdot k!\, S(r,k) \cdot \frac{\Gamma\left(k+\frac{3}{4}\right)\Gamma\left(\frac{5}{4}\right)}{\Gamma(k+2)} \\
&= \sum_{k=1}^{r}(-1)^k \cdot \frac{S(r,k)}{k+1} \cdot \Gamma\left(k+\tfrac{3}{4}\right)\Gamma\left(\tfrac{5}{4}\right) \\
&= \sum_{k=1}^{r}(-1)^k \cdot \frac{S(r,k)}{k+1} \cdot \left(\tfrac{3}{4}\right)\cdot\left(\tfrac{3}{4}+1\right)\cdots\left(\tfrac{3}{4}+k-1\right)\Gamma\left(\tfrac{3}{4}\right)\Gamma\left(\tfrac{5}{4}\right) \\
&= \sum_{k=1}^{r}(-1)^k \cdot \frac{S(r,k)}{k+1} \cdot \left(\tfrac{3}{4}\right)\cdot\left(\tfrac{3}{4}+1\right)\cdots\left(\tfrac{3}{4}+k-1\right)\tfrac{1}{2\sqrt{2}}\pi \\
&= \sum_{k=1}^{r}(-1)^k \cdot \frac{S(r,k)}{k+1} \cdot \left(\tfrac{3}{4}\right)^{(k)}\tfrac{1}{2\sqrt{2}}\pi,
\end{aligned}$$

where $\Gamma(\cdot)$ is the Gamma function.

But, from [Weisstein], we have

$$\frac{B_{r+1}\left(\frac{1}{4}\right)-B_{r+1}}{r+1}=\frac{(-2^{-(r+1)}(1-2^{-r})B_{r+1}-4^{-(r+1)}(r+1)E_r-B_{r+1})}{r+1}.$$

Thus, we have

$$B_{r+1}=-\frac{r+1}{4(1+2^{-(r+1)}(1-2^{-r}))}\left(\sum_{k=1}^{r}(-1)^k\cdot\frac{S(r,k)}{k+1}\cdot\left(\tfrac{3}{4}\right)^{(k)}+4^{-r}E_r\right).$$

□

If we let $r=2l$, an even integer, in (1) we immediately obtain:

Corollary 5. $$E_{2l}=-4^{2l}\sum_{k=1}^{2l}(-1)^k\cdot\frac{S(2l,k)}{k+1}\cdot\left(\tfrac{3}{4}\right)^{(k)}.$$

References

[Brychkov et al. 2019] Y. A. Brychkov, O. I. Marichev, and N. V. Savischenko, *Handbook of Mellin transforms*, CRC Press, Boca Raton, FL, 2019. MR Zbl

[Gould 1972] H. W. Gould, "Explicit formulas for Bernoulli numbers", *Amer. Math. Monthly* **79** (1972), 44–51. MR Zbl

[Jha 2019] S. K. Jha, "Two new explicit formulas for the Bernoulli numbers", preprint, 2019. arXiv

[Landsburg 2009] S. E. Landsburg, "Stirling numbers and polylogarithms", unpublished note, 2009, http://www.landsburg.com/query.pdf.

[Weisstein] E. W. Weisstein, "Bernoulli polynomial", http://mathworld.wolfram.com/BernoulliPolynomial.html. From MathWorld.

Received 17 Jun 2019. Revised 24 Jul 2019.

SUMIT KUMAR JHA:
kumarjha.sumit@research.iiit.ac.in
International Institute of Information Technology, Hyderabad, India

Moscow Journal of Combinatorics and Number Theory
Vol. 8, No. 4, 2019
dx.doi.org/10.2140/moscow.2019.8.389

Correction to the article
Intersection theorems for (0, ±1)-vectors and s-cross-intersecting families

Peter Frankl and Andrey Kupavskii

Volume 7:2 (2017), 91–109

We modify the statement and proof of Theorem 1 in "Intersection theorems for $\{0, \pm 1\}$-vectors and s-cross-intersecting families". A version of the paper that incorporates the errata is uploaded on the arXiv: https://arxiv.org/abs/1603.00938. We thank Danila Cherkashin and Sergei Kiselev for pointing out the error.

Part 2 of Theorem 1 in [Frankl and Kupavskii 2017] is incorrect. To give a corrected version, let us introduce some notation: $\mathcal{V}(n, m_1, m_2) \subset \{0, \pm 1\}^n$ is the collection of all vectors with exactly m_1 ones and m_2 minus ones, and

$$g(n, m_1, m_2) := \max\{|\mathcal{V}| : \mathcal{V} \subset \mathcal{V}(n, m_1, m_2) \text{ and } \langle \boldsymbol{v}, \boldsymbol{w}\rangle \geq -2m_2 + 1 \text{ for any } \boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}\}.$$

Theorem 1 (part 2). *For $n \geq n_0(k)$ and $0 \leq l \leq k$ we have*

$$F(n,k,-l) = \begin{cases} \sum_{i=0}^{l/2} \binom{k}{i}\binom{n}{k} & \text{for even } l, \\ g\big(n, k - \frac{l+1}{2}, \frac{l+1}{2}\big) + \sum_{i=0}^{(l-1)/2} \binom{k}{i}\binom{n}{k} & \text{for odd } l. \end{cases}$$

Thus, the statement is the same for even l and is different for odd l. The value of $g(n, m_1, m_2)$ seems to be very difficult to determine in general. We have studied this quantity in [Frankl and Kupavskii 2018a; 2018b]. In the former one, we determined the value of $g(n, m_1, 1)$ for any n, m_1. This allows us to determine exactly the value of $F(n, k, -1)$. In the latter one, we obtained the bounds

$$\binom{n}{k}\binom{k-1}{\frac{l-1}{2}} \leq g\big(n, k - \tfrac{l+1}{2}, \tfrac{l+1}{2}\big) \leq \binom{n}{k}\binom{k-1}{\frac{l-1}{2}} + \binom{n}{l+1}\binom{l+1}{\frac{l+1}{2}}\binom{n-l-2}{k-l-2}.$$

We go on to the proof. The proof is correct until (and including) Claim 3. Let us give the corrected version of the remainder of the proof. We first deal with the case of even l.

Claim 4. *Let l be even. If $I \subset \binom{[n]}{l+1}$ is bad, then for at least $\binom{n-k}{k-l-1}$ sets $S \supset I$, $S \in \binom{[n]}{k}$, we have $|\mathcal{V}_g(S)| \leq f(k,l) - 1$.*

Proof. Consider the family $\mathcal{A} \subset 2^S$ of subsets of S defined as $\mathcal{A} = \{N(\boldsymbol{w}) \cap S : \boldsymbol{w} \in \mathcal{V}_g(S)\}$. In view of the uniqueness part of Katona's theorem, it is sufficient to show that $\mathcal{A}$ does not contain one of the sets from the extremal family $\mathcal{U}^l$ for at least $\binom{n-k}{k-l-1}$ choices of S.

MSC2010: 05D05.
Keywords: families of vectors, intersecting family.

If I is bad then there exists a vector $\boldsymbol{v}$ of length $l+1$ such that both $\mathcal{V}(I,\boldsymbol{v})$ and $\mathcal{V}(I,\bar{\boldsymbol{v}})$ are nonempty. Assume without loss of generality that $|N(\boldsymbol{v})| \le l/2$ and take a vector $\boldsymbol{w} \in \mathcal{V}$ such that $\boldsymbol{w}|_I = \bar{\boldsymbol{v}}$. Then for any S such that $S \cap S(\boldsymbol{w}) = I$ the set $N(\boldsymbol{v})$ is missing from $\mathcal{A}$ (and, consequently, $|\mathcal{V}_g(S)| \le f(k,l)-1$). There are exactly $\binom{n-k}{k-l-1}$ such sets. □

Assume now that there are t bad sets $I \subset \binom{[n]}{l+1}$. Then the number of sets $S \subset \binom{[n]}{k}$ such that $|\mathcal{V}_g(S)| \le f(k,\ell)-1$ is at least $t\binom{n-k}{k-l-1}/\binom{k}{l+1}$. Therefore, by the original Claim 3 and the corrected Claim 4, we have

$$\begin{aligned}|\mathcal{V}| - f(k,l)\binom{n}{k} &\le -t\frac{\binom{n-k}{k-l-1}}{\binom{k}{l+1}} + \sum_{\text{bad } I}\frac{1}{2}\sum_{v\in\{\pm1\}^{l+1}}(|\mathcal{V}(I,\boldsymbol{v})| + |\mathcal{V}(I,\bar{\boldsymbol{v}})|)\\ &\le -t\left(\frac{\binom{n-k}{k-l-1}}{\binom{k}{l+1}} - 2^k(k-l)\binom{n-l-2}{k-l-2}\right) < 0,\end{aligned}$$

provided $n > 2^k k^2\binom{k}{l+1}$. We note that taking $n > 4^k k^2$ makes the choice of n for which the proof works independent of l.

The case of odd l turns out to be harder. We shall need the following variant of Katona's theorem.

Theorem I. *Assume that $\mathcal{F} \subset 2^{[n]} \setminus \binom{[n]}{m+1}$ and for any $F, G \in \mathcal{F}$ we have $|F \cup G| \le 2m+1$. Then $|\mathcal{F}| \le \sum_{i=0}^m \binom{n}{i}$; moreover, for $n > 2m+2$ the only example attaining the bound is $\bigcup_{i=0}^m \binom{[n]}{i}$.*

Proof. Without loss of generality, we may assume that $\mathcal{F}$ is shifted (we discuss the effect of this assumption on the uniqueness at the end of the proof). The proof is by induction. The statement is clear for $m=0$; moreover, the extremal family is unique. For $n = 2m+2$, it is easy to see that $2^{[n]} \setminus \binom{[n]}{m+1}$ splits into pairs of complementary sets, which implies the statement.

Assume that the statement holds for $(n-1,m)$ and $(n-1,m-1)$, and let us prove it for (n,m), $n > 2m+2$. We have $|\mathcal{F}| = |\mathcal{F}(n)| + |\mathcal{F}(\bar{n})|$. By induction, we have

$$|\mathcal{F}(\bar{n})| \le \sum_{i=0}^m \binom{n-1}{i}.$$

Moreover, by shiftedness it follows that $|F \cup G| \le 2m-1$ for any $F, G \in \mathcal{F}(n)$, and thus $|\mathcal{F}(n)| \le \sum_{i=0}^{m-1}\binom{n-1}{i}$. Since $n-1 > 2(m-1)+1$, this inequality is sharp unless $\mathcal{F}(n) = \bigcup_{i=0}^{m-1}\binom{[n-1]}{i}$. In the case of equality, from here it should be clear that $\mathcal{F}(\bar{n}) \subset \bigcup_{i=0}^m \binom{[n-1]}{i}$ and thus

$$\mathcal{F}(\bar{n}) = \bigcup_{i=0}^m \binom{[n-1]}{i}.$$

We remark that if $\mathcal{F}$ was not shifted initially, then it could not shift into $\bigcup_{i=0}^m \binom{[n]}{i}$; thus the uniqueness part holds for nonshifted families as well. □

Let us return to the case of odd l. Consider the subfamily $\mathcal{V}' \subset \mathcal{V}$ of all vectors from $\mathcal{V}$ that have exactly $\frac{l+1}{2}$ minus ones, and put $\mathcal{V}'' := \mathcal{V} \setminus \mathcal{V}'$. Arguing as in the case of even l, but applying Theorem I, we get

$$|\mathcal{V}''| \le \sum_{i=0}^{(l-1)/2}\binom{k}{i}\binom{n}{k},$$

and the inequality is sharp if $\mathcal{V}'' \neq \mathcal{U}$, where $\mathcal{U}$ consists of all $\{-1, 0, 1\}$-vectors with k nonzero coordinates and at most $\frac{l-1}{2}$ minus ones.

Note also that any vector with $\frac{l+1}{2}$ minus ones has scalar product at least $-l$ with any vector from $\mathcal{U}$. It is clear that $\mathcal{V}'$ must avoid scalar product $-l-1$. Moreover, it is sufficient for $\mathcal{V}' \cup \mathcal{U}$ to have all scalar products at least $-l$. Therefore, $|\mathcal{V}'| \leq g\bigl(n, k - \frac{l+1}{2}, \frac{l+1}{2}\bigr)$ and the largest $\mathcal{V}$ satisfying the requirements has size

$$g\bigl(n, k - \tfrac{l+1}{2}, \tfrac{l+1}{2}\bigr) + \sum_{i=0}^{(l-1)/2} \binom{k}{i}\binom{n}{k}.$$

References

[Frankl and Kupavskii 2017] P. Frankl and A. Kupavskii, “Intersection theorems for $\{0, \pm 1\}$-vectors and s-cross-intersecting families”, *Mosc. J. Comb. Number Theory* **7**:2 (2017), 91–109. MR Zbl

[Frankl and Kupavskii 2018a] P. Frankl and A. Kupavskii, “Erdős–Ko–Rado theorem for $\{0, \pm 1\}$-vectors”, *J. Combin. Theory Ser. A* **155** (2018), 157–179. MR Zbl

[Frankl and Kupavskii 2018b] P. Frankl and A. Kupavskii, “Families of vectors without antipodal pairs”, *Studia Sci. Math. Hungar.* **55**:2 (2018), 231–237. MR Zbl

Received 29 May 2019. Revised 16 Jul 2019.

PETER FRANKL:
peter.frankl@gmail.com
Rényi Institute, Hungarian Academy of Sciences, Budapest, Hungary

ANDREY KUPAVSKII:
kupavskii@ya.ru
Moscow Institute of Physics and Technology, Moscow, Russia
and
University of Oxford, Oxford, United Kingdom

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the submission page.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles are usually in English or French, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not refer to bibliography keys. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and a Mathematics Subject Classification for the article, and, for each author, affiliation (if appropriate) and email address.

Format. Authors are encouraged to use LaTeX and the standard amsart class, but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should normally be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages — Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc. — allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with as many details as you can about how your graphics were generated.

Bundle your figure files into a single archive (using zip, tar, rar or other format of your choice) and upload on the link you been provided at acceptance time. Each figure should be captioned and numbered so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text (“the curve looks like this:”). It is acceptable to submit a manuscript with all figures at the end, if their placement is specified in the text by means of comments such as “Place Figure 1 here”. The same considerations apply to tables.

White Space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal’s preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Moscow Journal of Combinatorics and Number Theory

vol. 8 no. 4 2019