Invent. math. 117, 27–56 (1994)

Inventiones mathematicae

On the period matrix of a Riemann surface of large genus (with an Appendix by J.H. Conway and N.J.A. Sloane)

P. Buser[1] **and P. Sarnak**[2]

[1] Département de Mathématiques, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne-Ecublens, Switzerland. e-mail: buser@dma.epfl.ch
[2] Department of Mathematics, Princeton University, Princeton NJ 08544, USA. e-mail: sarnak@math.princeton.edu

Oblatum 24-III-1993 & 25-VIII-1993

Summary. Riemann showed that a period matrix of a compact Riemann surface of genus $g \geqq 1$ satisfies certain relations. We give a further simple combinatorial property, related to the length of the shortest non-zero lattice vector, satisfied by such a period matrix, see (1.13). In particular, it is shown that for large genus the entire locus of Jacobians lies in a very small neighborhood of the boundary of the space of principally polarized abelian varieties.

We apply this to the problem of congruence subgroups of arithmetic lattices in $\mathrm{SL}_2(\mathbb{R})$. We show that, with the exception of a finite number of arithmetic lattices in $\mathrm{SL}_2(\mathbb{R})$, every such lattice has a subgroup of index at most 2 which is noncongruence. A notable exception is the modular group $SL_2(\mathbb{Z})$.

1 Introduction

Let P_n denote the set of positive $n \times n$ matrices of determinant 1. $\mathrm{SL}_n(\mathbb{Z})$ acts on P_n by

$$P \rightarrow \gamma P {}^t\gamma, \quad \gamma \in \mathrm{SL}_n(\mathbb{Z}) . \tag{1.1}$$

The quotient space $F_n = \mathrm{SL}_n(\mathbb{Z}) \backslash P_n$ parametrizes the space of flat n-dimensional (oriented) tori of unit volume (if $P = (p_{ij})$ then $ds^2 = p_{ij} dx_i dx_j$ on $\mathbb{Z}^n \backslash \mathbb{R}^n$ is such a flat torus).

Denote by $\mathfrak{h}_{2g}$ Siegel's space of symplectic P's

$$\mathfrak{h}_{2g} = \{ P \in P_{2g} \mid PJP = J \} \tag{1.2}$$

where $J = \left[\begin{smallmatrix} 0 & I_g \\ -I_g & 0 \end{smallmatrix}\right]$. Siegel's modular group $\mathrm{Sp}_{2g}(\mathbb{Z}) = \{ \gamma \in \mathrm{SL}_{2g}(\mathbb{Z}) \mid {}^t\gamma J \gamma = J \}$ acts on $\mathfrak{h}_{2g}$ by the action (1.1) and the quotient

$$\mathscr{A}_g = \mathrm{Sp}_{2g}(\mathbb{Z}) \backslash \mathfrak{h}_{2g} \tag{1.3}$$

parametrizes the principally polarized abelian varieties of (complex) dimension g, see [L, M]. (Usually the model for $\mathfrak{h}_{2g}$ is the set $\mathbb{H}^g$ of $Z = X + iY$, X, Y $g \times g$ real symmetrtic and $Y > 0$. The association $P = \left[\begin{smallmatrix} I & 0 \\ +X & I \end{smallmatrix}\right] \left[\begin{smallmatrix} Y^{-1} & 0 \\ 0 & Y^{+1} \end{smallmatrix}\right] \left[\begin{smallmatrix} I & +X \\ 0 & I \end{smallmatrix}\right]$ identifies this space with (1.2). See Appendix 0 for details and the relation to polarizations.)

Next we introduce the period sublocus of $\mathscr{A}_g$. Let $\mathscr{M}_g$ denote the moduli space of compact Riemann surfaces of genus g. Given $S \in \mathscr{M}_g$ and a canonical basis for the homology $H_1(S, \mathbb{Z})$, $\mathscr{N}_1, \ldots, \mathscr{N}_{2g}$, that is, the intersection matrix of the $\mathscr{N}_j$'s is J, we let $w_1, w_2, \ldots, w_{2g}$ be a dual basis of harmonic forms

$$\int_{\mathscr{N}_k} w_j = \delta_{jk} . \tag{1.4}$$

The corresponding period matrix P_S is defined to be (see A.0.5 for its relation to the usual period matrix)

$$P_S = (p_{ij}) = \left(\int_S w_i \wedge * w_j \right) . \tag{1.5}$$

Riemann's period relations [F-K] are equivalent to P_S being in $\mathfrak{h}_{2g}$. Note that another choice of a canonical basis for the homology yields, through the above association, the same point $P_S \in \mathscr{A}_g$. Thus, we have a map $t\colon \mathscr{M}_g \to \mathscr{A}_g$ which associates to a Riemann surface its period matrix. We will refer to the locus $t(\mathscr{M}_g)$ in $\mathscr{A}_g$ as the period locus or the locus of Jacobians. The Schottky problem is to describe this locus. Very interesting characterizations of $t(\mathscr{M}_g)$ in terms of properties of θ-divisors and in terms of $K-P$-equations have been found [Gu, Sh]. However, these do not lend themselves to an explicit or direct determination of whether a given $P \in \mathscr{A}_g$ is a period matrix (see, however, Farkas [F] for Schottky–Jung theory and in particular an explicit description of $t(\mathscr{M}_4)$).

We wish to study the location of $t(\mathscr{M}_g)$ in $\mathscr{A}_g$, this being more to the point in many applications. To do so, we investigate the range of the following function m on the various loci.

Define for $P \in P_n$

$$m(P) = \min_{\substack{n_1 \in \mathbb{Z}^n \\ n_1 \neq 0}} {}^t n_1 P n_1 . \tag{1.6}$$

m is clearly well defined on F_n and in fact is the square of the length of the shortest closed geodesic on the corresponding flat torus. By a well-known compactness theorem of Mahler [Gru-Le], $m(P) \to 0$ iff $P \to \partial(F_n)$. That is, $m = 0$ defines the boundary of F_n, and we may think of m as giving a "distance" function to $\partial(F_n)$. The inclusion

$$\mathscr{A}_g \subsetneq F_{2g} \tag{1.7}$$

is proper and finite (though not injective), so that the restriction of m to $\mathscr{A}_g$ gives a "distance" function to $\partial(\mathscr{A}_g)$. Define the constants $\gamma_{2g}, \delta_{2g}, \eta_{2g}$ by

$$\left.\begin{aligned} \gamma_{2g} &= \max_{P \in P_{2g}} m(P) \\ \delta_{2g} &= \max_{P \in \mathfrak{h}_{2g}} m(P) \\ \eta_{2g} &= \sup_{P \in t(\mathscr{M}_g)} m(P) \end{aligned}\right\} \tag{1.8}$$

So, including the supremum, the range of m on F_{2g} is $[0, \gamma_{2g}]$, on $\mathscr{A}_g$ it is $[0, \delta_{2g}]$ and on $t(\mathscr{M}_g)$ it is $[0, \eta_{2g}]$. The constant γ_{2g} is Hermite's constant from the geometry of numbers [Gru-Le]. Its size is fundamental in the theory of lattice packings since P's with large $m(P)$ give good packings of space by spheres.

Minkowski [Gru-Le] showed that

$$(\sigma_{2g}/2)^{-1/g} \leqq \gamma_{2g} \leqq 4\sigma_{2g}^{-1/g} \tag{1.9}$$

where

$$\sigma_n = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)}$$

is the volume of the unit n-ball in $\mathbb{R}^n$. For large g (1.9) gives, after an application of Stirling's series,

$$\frac{g}{\pi e} \leqq \gamma_{2g} \leqq \frac{4g}{\pi e}\,. \tag{1.10}$$

While the lower bound remains essentially the best that is known, the upper bound has been improved to

$$\gamma_{2g} \leqq \frac{1.744}{\pi e} g \tag{1.11}$$

in [K-L].

In Sect. 2 we show, using a variation on averaging methods from the geometry of numbers, that

$$(\sigma_{2g}/2)^{-1/g} \leqq \delta_{2g}\,. \tag{1.12}$$

Thus, essentially (as far as we know), m gets as large on $\mathscr{A}_g$ as it is on F_{2g}. In general δ_{2g} may not be equal to γ_{2g}, see (A.3). However, somewhat surprisingly, its behavior on the period locus is limited. We will prove

$$c\log g \leqq \eta_{2g} \leqq \frac{3}{\pi}\log(4g+3) \tag{1.13}$$

where c is a small positive constant. (See also (3.20) for a slight improvement of the upper bound.)

The upper bound (1.13) gives a strong combinatorial property that a period matrix satisfies. If we set

$$N_g = \left\{P \in \mathscr{A}_g \,|\, m(P) \leqq \frac{3}{\pi}\log(4g+3)\right\} \tag{1.14}$$

then N_g is a neighborhood of the boundary $\partial\mathscr{A}_g$ which contains the entire locus of Jacobians. We will show in Sect. 2 that if λ is the Riemannian volume form on $\mathscr{A}_g$ (coming from the symmetric space $\mathfrak{h}_{2g}$) normalized so that $\lambda(\mathscr{A}_g) = 1$, then

$$\lambda(N_g) = O(g^{-\nu g}) \quad \text{for any } \nu < 1\,. \tag{1.15}$$

So for g large N_g is very small (see Fig. 1). For g of the form 2^n, $n \geqq 7$, we show in Appendix 1 how (1.13) may be used to give explicit families of points in $\mathscr{A}_g$ which are not period matrices (see A.1.10). The smallest genus for which we have found (1.13) to be effective is $g = 12$. In Appendix 2 it is shown that the 24 dimensional Leech Lattice [C-S] is symplectic. It comes from an explicit point $Z_\Lambda \in \mathbb{H}^{12}$ and (1.13) shows Z_Λ is not a Jacobian. See Appendices 1 and 2 for details. In these appendices we also determine γ_{2g} and η_{2g} for small g.

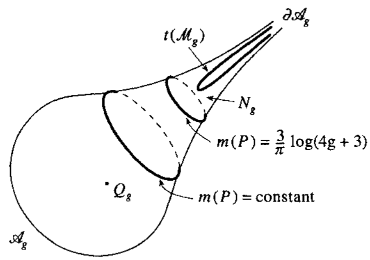

Fig. 1. A schematic picture of the locus $\mathscr{A}_g$ and $t(\mathscr{M}_g)$, Q_g corresponds to the Barnes–Wall lattice

The upper bound in (1.13) will be established in Sect. 3. We show that any $S \in \mathscr{M}_g$ carries a nonseparating cylinder of small capacity. This is used to construct a suitable harmonic form and a short lattice vector. The existence of the cylinder follows from (1.16), below which is of independent interest.

Let S be a compact Riemann surface of genus g equipped with its hyperbolic metric, then we can find on S a homologically non-trivial simple closed geodesic γ whose length is less than $2\log(4g-2)$ *and which has a collar of width greater than* $\operatorname{artanh}(2/3)$. (1.16)

By a collar in S about γ of width w we mean that the set of all points within a distance w from γ is a cylinder. The lower bound in (1.13) will be established in Sect. 4. Specific Riemann surfaces constructed by arithmetic means are shown to have $m(P_S)$ of oder $\log g$. To that end we use known lower bounds on the smallest eigenvalue of the Laplacian for congruence surfaces as well as the circle method from analytic number theory to construct surfaces of all genera.

We turn to the applications which concern arithmetic and congruence lattices in $SL_2(\mathbb{R})$. Firstly, we do not distinguish between a latticer Γ in $SL_2(\mathbb{R})$ or any of its conjugates in $SL_2(\mathbb{R})$. Arithmetic lattices are obtained as follows:

Let $k/\mathbb{Q}$ be a totally real number field with infinite places $\sigma_1, \sigma_2, \ldots, \sigma_n$. Let D be a quaternion algebra over k and assume that $D \otimes_{k,\sigma_1} \mathbb{R} \cong M_2(\mathbb{R})$ (by an isomorphism ϕ, say) and $D \otimes_{k,\sigma_j} \mathbb{R} \cong H(\mathbb{R})$ for $j = 2, \ldots, n$ where $H(\mathbb{R})$ is the Hamilton quaternions. Let D_1^* be the group of elements of D of reduced norm 1. If ρ is a faithful representation of D_1^* in $GL(n)$ defined over k, then a subgroup Φ of D_1^* is a congruence group if

$$\Phi^1 = \{x \in D_1^* \mid \rho(x) \in GL_n(\mathcal{O}_k),\ \rho(x) \equiv 1 (\operatorname{mod} \mathscr{A})\}$$

is containd in Φ with finite index, where $\mathcal{O}_k$ is the ring of integers of k and $\mathscr{A}$ is some ideal in $\mathcal{O}_k$. $\phi(\Phi)$ is then a congruence lattice in $SL_2(\mathbb{R})$. The arithmetic lattices in $SL_2(\mathbb{R})$ are lattices Γ for which $\Gamma \cap \Phi$ has finite index in both Γ and Φ for some congruence lattice Φ [Bo, Ra].

It is known that, given any arithmetic $\Gamma \leqq SL_2(\mathbb{R})$, there is a subgroup Γ' of Γ of finite index which is noncongruence. One way of seeing this is to use results on the congruence subgroup problem [B-M-S] which assert that if Γ has the congruence subgroup property (i.e. if the above failed) then $\Delta/[\Delta, \Delta]$ is finite for every finite index subgroup Δ of Γ. For lattices Γ in $SL_2(\mathbb{R})$ it is easily seen from their well-known structure in terms of generators and relations that there are subgroups

Δ of Γ of finite index for which $|\Delta/[\Delta, \Delta]| = \infty$. We show that for all but a finite number of arithmetic lattices in $\mathrm{SL}_2(\mathbb{R})$ such noncongruence subgroups exist at the smallest possible index. Precisely:

With the exception of a finite number of arithmetic lattices in $\mathrm{SL}_2(\mathbb{R})$, *every arithmetic lattice has a subgroup of index at most* 2 *which is noncongruence.* (1.17)

The proof will be given in Sect. 5. It proceeds by showing (using (1.13)) that all but finitely many arithmetic quotients have covers of degree at most two, with exceptional eigenvalues (see (5.3)).

This is in turn proven by first establishing the following (see Sect. 4 for the definition of λ_1).

Given $\varepsilon > 0$ *there is a* $g(\varepsilon)$ *such that if* Γ *is a lattice in* $\mathrm{SL}_2(\mathbb{R})$ *of genus* $g \geqq g(\varepsilon)$ *then* Γ *contains a subgroup* Γ^* *of index* 2 *such that* $\lambda_1(\Gamma^*) < \varepsilon$. (1.18)

The proof of this is given in Sect. 5, (5.3)–(5.9).

The deduction of (1.17) from (1.18) then follows from the finiteness theorem of Borel [Bo] on arithmetic groups with bounded volume, a general upper bound for $\lambda_1(\Gamma)$ of Zograf [Z] and known lower bounds for λ_1 for congruence groups.

Note that $\mathrm{SL}_2(\mathbb{Z})$ is one of the exceptions to (1.17). Its smallest index noncongruence subgroup is 7, see [Rn]. To end the introduction we draw the reader's attention to the recent preprint of Gromov [Gro] which puts the inequalities (1.13) and (1.16) in a general geometric setting and gives an account of a number of interesting related geometric inequalities.

2

We give two proofs of (1.12). The first is more elementary in that it does not appeal to the Haar measure on $\mathrm{Sp}_{2g}(\mathbb{R}) = \{g \in \mathrm{SL}_{2g}(\mathbb{R}) | {}^t gJg = J\}$. Let $U(\mathbb{R})$ be the subgroup of $\mathrm{Sp}_{2g}(\mathbb{R})$ defined as follows

$$U(\mathbb{R}) = \left\{ \begin{bmatrix} I & X \\ 0 & I \end{bmatrix} \middle| X \text{ is real } g \times g \text{ and symmetric} \right\}. \tag{2.1}$$

Let $U(\mathbb{Z})$ be the subgroup of $U(\mathbb{R})$ for which X is integral. A fundamental domain for $U(\mathbb{Z})$ in $U(\mathbb{R})$ is given by

$$\left\{ \begin{bmatrix} I & X \\ 0 & I \end{bmatrix} \middle| |X_{ij}| \leqq \frac{1}{2} \right\}. \tag{2.2}$$

Thus relative to the $U(\mathbb{R})$ invariant measure

$$dX = \prod_{i \leqq j} dX_{ij}, \quad \mathrm{Vol}(U(\mathbb{Z}) \backslash U(\mathbb{R})) = 1 . \tag{2.3}$$

Let f be an integrable function of compact support on $\mathbb{R}^{2g}$. Consider for $y > 0$ the function $I(y)$ defined by

$$I(y) = \int_{U(\mathbb{Z})\backslash U(\mathbb{R})} \sum_{\substack{m\in\mathbb{Z}^g \\ n\in\mathbb{Z}^g}}' f\left(\begin{bmatrix} y & & & & & \\ & \ddots & & & & \\ & & y & & & \\ & & & y^{-1} & & \\ & & & & \ddots & \\ & & & & & y^{-1} \end{bmatrix} \begin{bmatrix} I & X \\ 0 & I \end{bmatrix} \begin{bmatrix} {}^t m \\ {}^t n \end{bmatrix}\right) dX$$

$$= \int_{U(\mathbb{Z})\backslash U(\mathbb{R})} \sum_{m,n}' f\left(\begin{bmatrix} y & & & & & \\ & \ddots & & & & \\ & & y & & & \\ & & & y^{-1} & & \\ & & & & \ddots & \\ & & & & & y^{-1} \end{bmatrix}\right.$$

$$\times \left.\begin{bmatrix} m_1 + X_{11}n_1 + \ldots + X_{1g}n_g \\ \vdots \\ m_g + X_{g1}n_1 + \ldots + X_{gg}n_g \\ n_1 \\ \vdots \\ n_g \end{bmatrix}\right) dX \tag{2.4}$$

where $'$ denotes omit $(m, n) = (0, 0)$. If $n_1 \neq 0$ then, doing the X_{11} integral first and using the m_1 sum, we have

$$\sum_{m_1 \bmod n_1} \sum_{j=-\infty}^{\infty} \int_0^1 f\left(\begin{bmatrix} y n_1\left(\frac{m_1}{n_1} + j + X_{11} + \lambda\right) \\ \text{indep of } X_{11} \end{bmatrix}\right) dX_{11}$$

for λ independent of X_{11} and m_1

$$= n_1 \int_{-\infty}^{\infty} f\left(\begin{bmatrix} y n_1 X_{11} \\ \text{indep of } X_{11} \end{bmatrix}\right) dX_{11}$$

$$= y^{-1} \int_{-\infty}^{\infty} f\left(\begin{bmatrix} t_1 \\ y(m_2 + X_{21}n_1 + \cdots + X_{2g}n_g) \\ \vdots \end{bmatrix}\right) dt_1 .$$

Next do the X_{21} integral and m_2 sum and repeat the above argument. In this way we obtain from $n = (n_1, \ldots, n_g)$ with $n_1 \neq 0$ a contribution (after summing on m)

$$y^{-g} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f\left(\begin{bmatrix} t_1 \\ \vdots \\ t_g \\ y^{-1} n_1 \\ \vdots \\ y^{-1} n_g \end{bmatrix}\right) dt_1 \ldots dt_g \,.$$

If $n_1 = 0$ and $n_2 \neq 0$ then the above calculation yields a contribution of

$$y^{-g} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f\left(\begin{bmatrix} t_1 \\ \vdots \\ t_g \\ 0 \\ y^{-1} n_2 \\ \vdots \\ y^{-1} n_g \end{bmatrix}\right) dt_1 \ldots dt_g \,.$$

Continuing in this way, we get finally from the $n = 0$ term

$$\sum_{m \neq 0} f\left(\begin{bmatrix} y^t m \\ 0 \end{bmatrix}\right).$$

Collecting these together, we have

$$I(y) = \sum_{n \neq 0} y^{-g} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, \ldots, t_g; y^{-1} n_1, \ldots, y^{-1} n_g)\, dt_1 \ldots dt_g + \sum_{m \neq 0} f(ym, 0) \,. \tag{2.5}$$

Since f is of compact support, we find on letting $y \to \infty$ that $I(y) \to I(\infty)$, where

$$I(\infty) = \int_{\mathbb{R}^{2g}} f(x)\, dx \,. \tag{2.6}$$

Now let

$$f(x) = \chi_{R^2}({}^t x x) \tag{2.7}$$

where χ_{R^2} is the characteristic function of $(0, R^2]$. Then

$$I(y) = \int_{U(\mathbb{Z}) \setminus U(\mathbb{R})} \sum_{\ell \in \mathbb{Z}^{2g}}{}' \chi_{R^2}({}^t \ell P_{X,y} \ell)\, dX \tag{2.8}$$

where

$$P_{X,y} = \begin{bmatrix} I & 0 \\ X & I \end{bmatrix} \begin{bmatrix} y^2 I & 0 \\ 0 & y^{-2} I \end{bmatrix} \begin{bmatrix} I & X \\ 0 & I \end{bmatrix}. \tag{2.9}$$

Also

$$\int_{\mathbb{R}^{2g}} \chi_{R^2}({}^t x x)\, dx = \int_{\|x\| \leq R} dx = R^{2g} \sigma_{2g} \,. \tag{2.10}$$

Now if $R^{2g}\sigma_{2g} < 2$, then

$$I(\infty) < 2 \,. \tag{2.11}$$

Hence for some y and X

$$\sum_{\ell\in\mathbb{Z}^{2g}}{}' \chi_{R^2}({}^t\ell P_{X,y}\ell) < 2 \,.$$

But since the last sum clearly takes only even integral values, it follows that

$$\sum_{\ell\in\mathbb{Z}^{2g}}{}' \chi_{R^2}({}^t\ell P_{X,y}\ell) = 0$$

and therefore $m(P_{X,y}) \geqq R^2$. From (2.11) and $P_{X,y}\in\mathscr{A}_g$, we conclude that

$$\delta_{2g} \geqq \left(\frac{2}{\sigma_{2g}}\right)^{1/g}$$

proving (1.12).

The second method uses a $G = \mathrm{Sp}_{2g}(\mathbb{R})$ invariant measure $\tilde{\lambda}$ on $\mathrm{Sp}_{2g}(\mathbb{Z})\backslash\mathrm{Sp}_{2g}(\mathbb{R})$. This measure descends to one on $\mathrm{Sp}_{2g}(\mathbb{Z})\backslash\mathfrak{h}_{2g} = \mathscr{A}_g$ which we denote by $\tilde{\lambda}$ as well. If one follows the computation of the volume $\tilde{\lambda}(\mathscr{A}_g)$ in Siegel [Si, pp. 325–330] and uses his normalization of $\tilde{\lambda}$, one finds that he proves the following:

If F is an integrable function of compact support on $[0,\infty)$, then

$$\int_{\mathscr{A}_g}\sum_{\ell\in\mathbb{Z}^{2g}}{}' F({}^t\ell P\ell)\,d\tilde{\lambda}(P) = \zeta(2g)\,V_{g-1}\int_0^\infty F(x)x^{g-1}\,dx \tag{2.12}$$

where

$$V_g = \int_{\mathscr{A}_g} d\tilde{\lambda}(P)$$

and ζ is the Riemann zeta function. He also shows that

$$V_g = (g-1)!\,\pi^{-g}\zeta(2g)\,V_{g-1} \,. \tag{2.13}$$

Thus, if λ is the unique invariant measure on $\mathscr{A}_g$ normalized so that

$$\lambda(\mathscr{A}_g) = 1 \,, \tag{2.14}$$

then (2.12) reads

$$\int_{\mathscr{A}_g}\sum_{\ell\in\mathbb{Z}^{2g}}{}' F({}^t\ell P\ell)\,d\lambda(P) = \frac{\pi^g}{(g-1)!}\int_0^\infty F(x)x^{g-1}\,dx \,. \tag{2.15}$$

If F_R is chosen to be χ_{R^2} as in (2.7), this gives

$$\int_{\mathscr{A}_g}\sum_{\ell\in\mathbb{Z}^{2g}}{}' \chi_{R^2}({}^t\ell P\ell)\,d\lambda(P) = \frac{R^{2g}\pi^g}{g!} = R^{2g}\sigma_{2g} \,. \tag{2.16}$$

Comparing this with (2.10) shows that this implies the same lower bound for δ_{2g} as method 1. Actually, the reason they give the same answer is that the measures μ_g on $\mathrm{Sp}_{2g}(\mathbb{Z})\backslash\mathrm{Sp}_{2g}(\mathbb{R})$ defined by

$$\mu_g(f) = \int_{U(\mathbb{Z})\backslash U(\mathbb{R})} f\left(\begin{bmatrix} yI & 0 \\ 0 & y^{-1}I\end{bmatrix}\begin{bmatrix} I & X \\ 0 & I\end{bmatrix}\right)dX$$

tend weakly to λ as $y\to\infty$, but this will not concern us here.

To prove (1.15), note that for $P \in N_g$ (where N_g is defined by (1.14)),

$$\sum_{\ell \in \mathbb{Z}^{2g}}' \chi_{R_0^2}({}^t\ell P \ell) \geqq 2$$

where

$$R_0^2 = \frac{3}{\pi} \log(4g + 3) .$$

Hence

$$\begin{aligned} \lambda(N_g) &\leqq \int_{N_g} \sum_{\ell \in \mathbb{Z}^{2g}}' \chi_{R_0^2}({}^t\ell P \ell)\, d\lambda(P) \\ &\leqq \int_{\mathscr{A}_g} \sum_{\ell \in \mathbb{Z}^{2g}}' \chi_{R_0^2}({}^t\ell P \ell)\, d\lambda(P) \\ &= R_0^{2g} \sigma_{2g} \quad \text{by (2.16).} \end{aligned}$$

Hence $\lambda(N_g) = O(g^{-\alpha g})$ for any $\alpha < 1$. This proves (1.15).

3

We begin by giving a more useful characterization of $m(P_S)$ for $S \in \mathscr{M}_g$. We claim that

$$m(P_S) = \inf \int_S w \wedge * w \tag{3.1}$$

where the inf is taken over all closed (real) 1-forms w on S which have integral periods over all cycles and not all periods are zero.

To see this first note that the inf for a given set of periods is assumed by a harmonic w as is well known. If $\mathscr{N}_1, \ldots, \mathscr{N}_{2g}$ is a canonical basis for the homology and $w_1, \ldots, w_{2g}$ a dual basis, then our form w is of the form

$$w = m_1 w_1 + \cdots + m_{2g} w_{2g} .$$

Now by the definition of the w's,

$$m_j = \int_{\mathscr{N}_j} w$$

and since all periods of w are in $\mathbb{Z}$, $m_j \in \mathbb{Z}$. Also, $m = (m_1, \ldots, m_{2g}) \neq 0$. From the definition of P_S, it follows that

$$\int_S w \wedge * w = {}^t m P_S m .$$

Thus the infimum in (3.1) satisfies

$$\inf \int_S w \wedge * w \geqq m(P_S) .$$

Conversely, if $m(P_S)$ is achieved by $0 \neq \hat{m} = (\hat{m}_1, \ldots, \hat{m}_{2g})$, then set

$$\tilde{w} = \hat{m}_1 w_1 + \cdots + \hat{m}_{2g} w_{2g} .$$

$\tilde{w}$ has integral periods (not all zero) and hence

$$\inf \int_S w \wedge * w \leqq \int_S \tilde{w} \wedge * \tilde{w} = m(P_S) .$$

This proves (3.1).

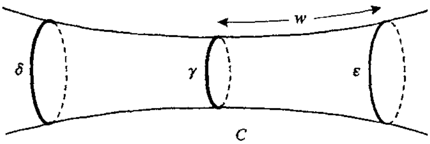

Fig. 2

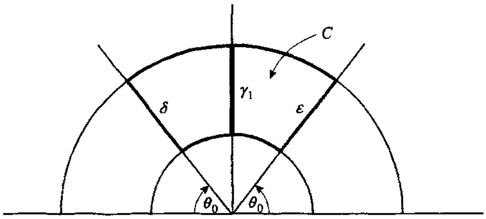

Fig. 3

To prove the upper bound in (1.13), we use (3.1) by finding on any $S \in \mathcal{M}_g$ a suitable test form w. In fact, suppose we have equipped S with its hyperbolic metric and that we have a simple closed geodesic γ which is homologically nontrivial. Let γ be of length ℓ and suppose it has a collar C (in S) of width w. Since γ is homologically nontrivial, it does not separate S into two pieces.

We construct a suitable test form w as follows. Let δ and ε be the boundary components of C and let F be defined on C with $F|_\delta = 1$, $F|_\varepsilon = 0$. Set $w = dF$ on C and $w = 0$ on $S \setminus C$. Then w is a suitable test form in (3.1), for it clearly has integral periods, and since γ does not separate, not all its periods are zero. In fact, some periods equal one. Now

$$\int_S w \wedge *w = \int_C dF \wedge (*dF) .$$

The minimum of the last expression subject to $F|_\delta = 1$ and $F|_\varepsilon = 0$ is simply the capacity cap(C) of the annulus (cylinder) C. Thus

$$m(P_S) \leqq \operatorname{cap}(C) . \tag{3.2}$$

To compute the capacity, we use the hyperbolic plane model for C. The notation is as in Fig. 3. The semicircles have a common perpendicular γ_1 of length $\ell(\gamma_1) = \ell(\gamma) = \ell$. They are identified by $z \mapsto \lambda z$.

We have

$$\left.\begin{aligned} &\log \lambda = \ell \\ &\text{and} \\ &w = \int_{\theta_0}^{\pi/2} \frac{d\theta}{\sin\theta} \end{aligned}\right\} \tag{3.3}$$

The mapping $z \mapsto w = e^{(2\pi i \log z)/\log \lambda}$ gives a conformal mapping of C onto the annulus $A = \{R_1 \leqq |z| \leqq R_2\}$ where

$$R_1 = e^{-2\pi(\pi - \theta_0)/\log \lambda}, \quad R_2 = e^{-2\pi\theta_0/\log \lambda}.$$

Hence

$$\operatorname{cap}(C) = \operatorname{cap}(A) = 2\pi \frac{\log \lambda}{2\pi(\pi - 2\theta_0)}. \tag{3.4}$$

Thus

$$m(P_S) \leqq \frac{\ell}{(\pi - 2\theta_0)} \tag{3.5}$$

where θ_0 is determined by (3.3) or equivalently by

$$\cosh w = 1/\sin \theta_0 . \tag{3.6}$$

Next we prove (1.16) which in connection with (3.5) leads to the bound

$$m(P_S) \leqq \frac{2\log(4g-2)}{(\pi - 2\theta_0)}$$

where $\theta_0 = \arccos(2/3)$ and so

$$m(P_S) \leqq (1.37\ldots)\log(4g-2).$$

An additional distinction of cases (see (3.12)) will improve the bound as in (1.13).

The proof of (1.16) is in two steps. First, we show that there exists on S a non-separating simple closed geodesic of length less than L, where

$$L = 2\log(4g-2).$$

This step will use area arguments. In the second step we use hyperbolic trigonometry to show that the shortest non-separating simple closed geodesic has a collar of width at least artanh(2/3).

Let $q \in S$ and denote by r_q the injectivity radius of S at q. This is the supremum of all r such that the distance set

$$D = \{x \in S \mid \operatorname{dist}(x, q) \leqq r\}$$

is isometric to a disk of radius r in the hyperbolic plane. Since D has area $2\pi(\cosh r_q - 1)$ and, by Gauss–Bonnet, S has area $4\pi(g-1)$, we obtain

$$r_q < \log(4g-2). \tag{3.7}$$

By the definition of r_q there exist two geodesic arcs of length r_q emanating from q and meeting at their endpoints. Since r_q is minimal with this property, the two arcs together form a *geodesic loop* c at q, that is, a (smooth) geodesic arc whose endpoints coincide with q. The existence of c may also be seen in the universal covering $\mathbb{H}$ of S by considering two distinct lifts of q at minimal distance $(2r_q)$ from each other. Since no lift of c in $\mathbb{H}$ is a closed curve, c is a homotopically nontrivial closed curve. We let γ be the closed geodesic in the free homotopy class of c. Since c is simple, γ is simple. We refer to [F-L-P] or [Bu] for this and for related intersection properties of closed geodesics which will be tacitly used in the sequel.

So far, we have found a simple closed geodesic γ on S of length less than L. If γ is non-separating, we are done. Let us now assume that γ separates S into two bordered surfaces F and F', each having a copy of γ as boundary. We let F be the component with the smaller area. Then F is a compact Riemann surface of signature $(h; 1)$ with $h \leqq g/2$, and we have the following general fact.

Let F be a compact Riemann surface of signature $(h; 1)$ with $1 \leqq h \leqq \frac{1}{2}g$ and assume that the boundary γ has length $\ell(\gamma) < L = 2\log(4g-2)$. Then F contains another simple closed geodesic of length less than L in its interior. (3.8)

With (3.8) we achieve our first step. For if γ on F is non-separating we are done. But if γ separates, then we cut F open along γ to obtain two pieces, one of area $\geqq 2\pi$ (by Gauss–Bonnet) and the other, F' say, again of signature $(1; h')$ for some h'. We then apply (3.8) to F' finding another geodesic of length less than L, and so on. Eventually, one of these geodesics is non-separating.

Proof of (3.8) We consider the boundary collar

$$C_\gamma = \{x \in F \mid \operatorname{dist}(x, \gamma) < w\}$$

where w is the supremum of all ω such that the geodesic arcs of length ω emanating perpendicularly from γ are pairwise disjoint. There exist two geodesic arcs of length w perpendicular to γ having their endpoint p in common. Since w is minimal with this property, the two arcs together form a smooth geodesic arc δ as shown in Fig. 4.

Again, the existence of δ may also be seen in the universal covering by looking at two distinct lifts of γ at minimal distance $(2w)$ from each other.

The endpoints of δ separate γ into two arcs γ_1 and γ_2 with $\ell(\gamma_1) \leqq \frac{1}{2}\ell(\gamma)$. The closed curve $\gamma_1\delta$ (first along γ_1 then along δ) is homotopically nontrivial, for otherwise it would have a closed lift in the universal covering consisting of two perpendicular geodesic arcs which is impossible. Similarly, $\delta^{-1}\gamma_2$ is homotopically nontrivial. Since neither $\gamma_1\delta$ nor $\delta^{-1}\gamma_2$ are contractible, the curve $\gamma_1\delta$ is not freely homotopic to the boundary γ (or γ^{-1}) of F. Indeed if $\gamma_1\delta$ and γ are freely homotopic and γ is a boundary curve, then by a well known fact in surface topology [Bu, Proposition A.11; E] $\gamma_1\delta$ and γ bound a domain of signature $(0; 2)$ and it follows that $\delta^{-1}\gamma_2$ bounds a disk.

Now let β be the simple closed geodesic in the free homotopy class of $\gamma_1\delta$. Then β is contained in the interior of F and satisfies

$$\ell(\beta) < 2w + \tfrac{1}{2}\ell(\gamma) .$$

If $2w + \ell(\gamma)/2 \leqq L$, then β is as in (3.8). ((3.8) does not require β to be non-separating.) We assume, therefore, from now on that

$$2w + \tfrac{1}{2}\ell(\gamma) > L . \tag{3.9}$$

Consider the double $\tilde{F}$ of F (that is, attach a mirror image of F along the boundary γ). In $\tilde{F}$, γ has an open collar $\tilde{C}_\gamma$ of width w,

$$\tilde{C}_\gamma = \{x \in \tilde{F} \mid \operatorname{dist}(x, \gamma) < w\} .$$

Let λ be the shortest geodesic loop at the midpoint p of δ. By (3.7),

$$\ell(\lambda) = 2r_p < L .$$

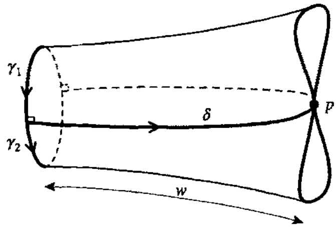

Fig. 4

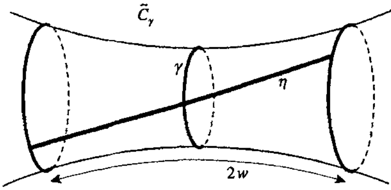

Fig. 5

We shall prove that

λ is contained in F ; (3.10)

λ is not freely homotopic to the boundary of F . (3.11)

The closed geodesic in the free homotopy class of λ will then satisfy the requirements of (3.8).

Proof of (3.10) Assume that λ intersects γ.

Then λ contains two subarcs such as arc η in Fig. 5 connecting the two boundary curves of $\tilde{C}_\gamma$ with each other (recall that γ is a separating geodesic in $\tilde{F}$). In view of (3.9) and the hypothesis of (3.8) ($\ell(\gamma) < L$), we get

$$\ell(\lambda) \geqq 4w > 2L - \ell(\gamma) > L ,$$

a contradiction.

Proof of (3.11) Let Λ be any geodesic loop at p in F freely homotopic to γ (or γ^{-1}). Λ is homotopic with fixed basepoint to $a\gamma a^{-1}$, where a is a geodesic arc from p to γ arriving perpendicularly at γ. The shortest possible arc of this kind is the arc a_1 of length $\ell(a_1) = \frac{1}{2}\ell(\delta) = w$.

Figure 6 shows lifts in the universal covering of the curves $a\gamma a^{-1}$ and $a_1\gamma a_1^{-1}$. The geodesic arc $\tilde{\Lambda}$ from $\tilde{p}$ to p' is a lift of Λ, and the geodesic arc $\tilde{\Lambda}_1$ is mapped under the universal covering map onto a geodesic loop Λ_1 at p. It is well known that the length of $\tilde{\Lambda}$ is a monotone increasing function of $\ell(\tilde{a})$ (the formula is $\sinh(\frac{1}{2}\ell(\tilde{\Lambda})) = \sinh(\frac{1}{2}\ell(\tilde{\gamma}))\cdot\cosh(\frac{1}{2}\ell(\tilde{a}))$, [Be, p. 157]), therefore $\ell(\Lambda) \geqq \ell(\Lambda_1)$.

Figure 7 shows a lift of the closed curve $a_1\gamma_1 a_2^{-1}$ where a_2 is the second half of δ. The arc $\tilde{\Lambda}_2$ is mapped onto a geodesic loop Λ_2 at p. Since $\ell(\tilde{\Lambda}_2)$ is a strictly

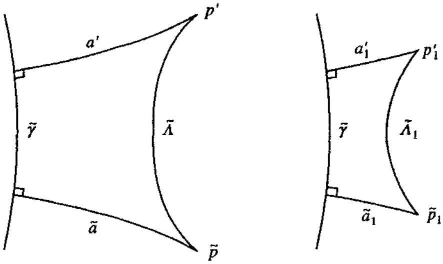

Fig. 6

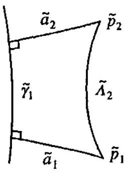

Fig. 7

monotone increasing function of $\ell(\gamma_1)$, we obtain altogether

$$\ell(A_2) < \ell(A_1) \leqq \ell(A) .$$

This shows that, if A is freely homotopic to γ, then A is not the shortest loop at p. (3.11) and (3.8) are now proven.

We turn to the second step. The constant μ occurring in the next statement is determined by the equation

$$\cosh\left(\frac{\mu}{2}\right) = 2\cosh\left(\frac{\mu}{4}\right).$$

The numerical value is $\mu = 3.325\ldots$.

The shortest non-separating simple closed geodesic γ on S has a collar of width

$$w > \operatorname{artanh}\tfrac{2}{3} = 0.8047\ldots.$$

This bound is sharp for any $g \geqq 2$. If $\ell(\gamma) > \mu$ then S has a collar of width

$$w > \operatorname{artanh} 2 = 1.316\ldots.$$

This bound is asymptotically sharp as $\ell(\gamma) \to \infty$. (3.12)

Proof. There exists a geodesic arc δ of length $2w$ meeting γ perpendicularly at its endpoints. Two cases are possible: either δ arrives at γ on opposite sides of γ or it arrives on the same side.

Case 1 δ arrives on opposite sides

Figure 8 shows a lift $\tilde{\delta}$ of δ in the universal covering, together with lifts $\tilde{\gamma}$ and γ^* of γ at the endpoints of $\tilde{\delta}$. Note that $\tilde{\gamma}$ and γ^* have the same orientations with

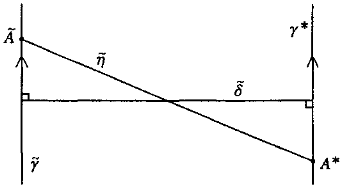

Fig. 8

respect to $\tilde{\delta}$. Therefore, there are points $\tilde{A} \in \tilde{\gamma}$ and $A^* \in \gamma^*$ on opposite sides of $\tilde{\delta}$ and at the same distance from $\tilde{\delta}$ such that $\tilde{A}$ and A^* are mapped to the same point $A \in \gamma$ under the universal covering map.

Observe that we can find $\tilde{A}$ and A^* in such a way that their distance to $\tilde{\delta}$ is at most $\frac{1}{4}\ell(\gamma)$. By drawing the geodesic $\tilde{\eta}$ from $\tilde{A}$ to A^*, we obtain two isometric right-angled geodesic triangles. Applying the cosine formula (Beardon [Be, p. 146]) to these triangles, we get

$$\cosh\tfrac{1}{4}\ell(\gamma)\cdot\cosh\,\ell(\tilde{\delta}) \geqq \cosh\tfrac{1}{2}\ell(\tilde{\eta})\,. \tag{3.13}$$

Since $\tilde{\eta}$ intersects $\tilde{\gamma}$ and γ^* under the same angle, the image η of $\tilde{\eta}$ under the universal covering map is a smooth closed geodesic on S. The curve γ contains an arc connecting points on opposite sides of η without intersecting η. Hence, η is non-separating and therefore $\ell(\eta) \geqq \ell(\gamma)$. Since $w = \frac{1}{2}\ell(\delta) = \frac{1}{2}\ell(\tilde{\delta})$, the above inequality implies

$$\cosh\tfrac{1}{4}\ell(\gamma)\cdot\cosh w \geqq \cosh\tfrac{1}{2}\ell(\gamma)\,. \tag{3.14}$$

The resulting lower bound for w is good for large $\ell(\gamma)$ but tends to 0 as $\ell(\gamma) \to 0$. We apply, therefore, also Randol's inequality $\sinh\frac{1}{2}\ell(\gamma)\cdot\sinh w > 1$ (Randol [R2]; this inequality holds for any simple closed geodesic on S). Both bounds together yield

$$w \geqq \max\left\{\operatorname{arcosh}\frac{\cosh\frac{1}{2}\ell(\gamma)}{\cosh\frac{1}{4}\ell(\gamma)};\ \operatorname{arsinh}\frac{1}{\sinh\frac{1}{2}\ell(\gamma)}\right\}. \tag{3.15}$$

Fom the monotonicity of the two functions in the brackets, we see that max has a strict global minimum w_0. The functions coincide when $\sinh(\ell(\gamma)/4) = \frac{1}{2}$ and so $\tanh w_0 = \frac{2}{3}$. This proves the first inequality; the second inequality follows from the definition of μ and the monotonicity of the first function in the brackets.

Case 2 The endpoints of δ meet γ on the same side of γ

This case is shown in Fig. 9. The endpoints of δ divide γ into two arcs γ_1 and γ_2. We denote by γ' and γ'' the closed geodesics in the free homotopy classes of the closed curves $\gamma_1\delta$ and $\delta^{-1}\gamma_2$. Since a small distance neighborhood of the union set $\gamma \cup \delta$ is a topological surface of signature (0; 3), it follows from Epstein's theorem on isotopies [E] that γ, γ' and γ'' bound a three-holed sphere Y (the interior of Y is embedded in S but some of the boundary geodesics may coincide in S).

It is well known (Fathi et al. [F-L-P], Thurston [T]) that the common perpendiculars (dotted lines in Fig. 9) dissect Y into right-angled geodesic hexagons and that δ decomposes these hexagons into right-angled pentagons.

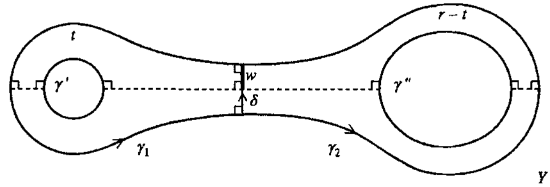

Fig. 9

Since γ is non-separating, one of the other boundary geodesics of Y, say γ'', is also a non-separating geodesic of S and so $\ell(\gamma'') \geqq \ell(\gamma)$.

We set $t = \frac{1}{2}\ell(\gamma_1)$, $r = \frac{1}{2}\ell(\gamma)$ and $r' = \frac{1}{2}\ell(\gamma')$. By the pentagon formula (Beardon [Be, p. 159]) we have

$$\sinh w \cdot \sinh t = \cosh r'$$

$$\sinh w \cdot \sinh(r - t) = \cosh \tfrac{1}{2}\ell(\gamma'') \geqq \cosh r\ . \tag{3.16}$$

Using the rule $\sinh(r - t) = \sinh r \cdot \cosh t - \cosh r \cdot \sinh t$, we get

$$\sinh w \cdot \cosh t \geqq \coth r \cdot (1 + \cosh r') \geqq 1 + \cosh r'\ .$$

Squaring the last line and using the first Eq. (3.16) once more, we get

$$\sinh^2 w + \cosh^2 r' \geqq 1 + 2\cosh r' + \cosh^2 r' > 3 + \cosh^2 r'\ .$$

This proves the inequality (3.12) in the second case. Together with (3.4), (3.6) and using that $\log \lambda = \ell = \ell(\gamma) < 2\log(4g - 2)$ we obtain the following result which is of independent interest.

The shortest non-separating simple closed geodesic γ on S has a collar of capacity $\mathrm{cap}(C) \leqq \frac{3}{\pi}\log(4g + 3)$. (3.17)

Together with (3.2) we get the upper bound in (1.13).

In order to prove that the bound $\operatorname{artanh}\frac{2}{3}$ in (3.12) is sharp we consider a 3-holed sphere Y with the boundary geodesics γ, γ'' and γ' of lengths $\ell(\gamma) = \ell(\gamma'') = \ell_0$ and $\ell(\gamma') = \varepsilon$ where ε is arbitrarily small and ℓ_0 satisfies $\sinh \ell_0/4 = \frac{1}{2}$.

We let δ be the shortest arc on Y from γ to γ''. Similar arcs exist between γ and γ' and between γ'' and γ'. The three arcs together decompose Y into two isomeric hexagons. By dropping the common perpendicular from δ to γ' on one of these hexagons we obtain two isometric right angled geodesic pentagons with sides of lengths $\frac{1}{2}\ell(\delta), \cdot, \frac{1}{4}\varepsilon, \cdot, \frac{1}{2}\ell(\gamma)$. The pentagon formula yields $\sinh\frac{1}{2}\ell(\gamma) \cdot \sinh\frac{1}{2}\ell(\delta) = \cosh\frac{\varepsilon}{4}$. As $\varepsilon \to 0$, the length of δ tends towards twice the bound w_0 of the width of a collar associated with a geodesic of length ℓ_0.

Now we identify γ with γ'' so that Y turns into a surface Q of signature (1; 1). When gluing γ to γ'' there is a degree of freedom: the Fenchel–Nielsen twist parameter. We choose this parameter such that the endpoints of δ in Q become opposite points on γ. The geodesic η occurring in the proof (case 1) has length $\ell(\eta)$ satisfying $\cosh\frac{1}{2}\ell(\eta) = \cosh\frac{1}{4}\ell_0 \cdot \cosh\frac{1}{2}\ell(\delta)$. This shows that γ is smaller than η and, in fact, γ is the shortest closed geodesic in the interior of Q. As $\varepsilon \to 0$, $\ell(\eta) \to \ell(\gamma)$. Since the width of the collar around γ must be less than $\frac{1}{2}\ell(\delta)$, we

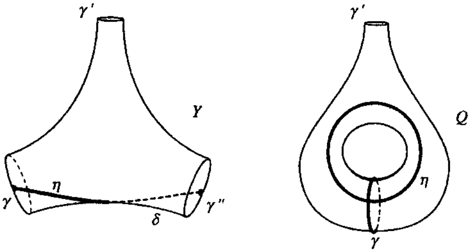

Fig. 10

obtain the sharpness of the first inequality in (3.12) for surfaces of signature (1; 1). Using a hyperbolic surface B of signature (0; g) with g boundary geodesics of length ε and attaching g copies of Q to B, we obtain sharpness for any genus $g \geqq 2$.

For the asymptotic sharpness of the second inequality in (3.12) we construct examples in a different spirit. We let $r > \mu/2$ be an arbitrarily large number and consider Y with the notation as in Fig. 9 (where δ goes from γ back to γ). We take $\ell(\gamma) = \ell(\gamma'') = 2r$. Now (3.16) tells us that $\cosh \frac{1}{2}\ell(\delta) \to 2$ as $r \to \infty$ and $\varepsilon \to 0$.

Now we take h copies of Y, and identify γ'' of the first copy with γ of the second, then γ'' of the second with γ of the third, and so on. Finally we identify γ'' of the last copy with γ of the first again. In this way we obtain a hyperbolic surface T of signature (1; h). By taking h large enough, the copies of γ on T are the shortest non-separating closed geodesics on T. This shows the asymptotic sharpness of the second inequality in (3.12) for surfaces such as T. To obtain examples without boundary we attach to each boundary geodesic of T a copy of a hyperbolic surface N where N has exactly one boundary geodesic (of length ε) and where the smallest closed geodesic in the interior of N is longer than r. That surfaces N with these properties exist follows from the next section.

For the upper bound in (1.13) we used the capacity of cylinders. Using more general domains we may improve the bound a little bit. We briefly outline the idea.

For γ an arbitrary non-separating closed geodesic on S we let W^* be the supremum of all numbers $\tilde{W}$ for which the set

$$\tilde{C} = \{x \in S \mid \operatorname{dist}(x, \gamma) < \tilde{W}\}$$

is separated by γ (that is, γ is a separating curve on $\tilde{C}$). Note that W^* is larger or equal to the width of the largest collar about γ.

Now take $\tilde{W} < W^*$ arbitrarily close to W^*. Denote by δ the union of the boundary components of $\tilde{C}$ on one side of γ and by ε the union of the boundary components on the other side. Define the capacity of $\tilde{C}$ with respect to γ as

$$\operatorname{cap}_\gamma(\tilde{C}) = \inf \int_{\tilde{C}} dF \wedge (*dF)$$

for test functions with $F \mid \delta = 1$, $F \mid \varepsilon = 0$. It is not difficult to prove that

$$m(P_S) \leqq \operatorname{cap}_\gamma(\tilde{C}) \leqq \operatorname{cap}(C_{\tilde{W}}) \tag{3.18}$$

where $C_{\tilde{W}}$ is a hyperbolic comparison cylinder (not contained in S) with core geodesic of length $\ell(\gamma)$ and width $\tilde{W}$.

Now let again γ be the shortest non-separating closed geodesic on S. To find W^* in this case one proceeds as in (3.12) where now only case 1 must be considered. The result is

$$\cosh W^* \geqq \frac{\cosh \frac{1}{2}\ell(\gamma)}{\cosh \frac{1}{4}\ell(\gamma)} . \tag{3.19}$$

Applying (3.4) and (3.6) to $C_{\tilde{W}}$ and using that $\tilde{W}$ may be chosen arbitrarily close to W^*, we obtain from (3.18)

$$m(P_S) \leqq \frac{\ell(\gamma)}{\pi - 2\arcsin\left(\dfrac{\cosh\frac{1}{4}\ell(\gamma)}{\cosh\frac{1}{2}\ell(\gamma)}\right)} \tag{3.20}$$

where, by (1.16), $\ell(\gamma) \leqq 2\log(4g-2)$.

4

In order to prove the lower bound in (1.13), we need to introduce the notion of $\lambda_1(\Gamma)$ for a lattice Γ in $\mathrm{SL}_2(\mathbb{R})$. Let $\mathbb{H}$ denote the hyperbolic plane and define

$$\lambda_1(\Gamma) = \inf \frac{\int\limits_{\Gamma\backslash\mathbb{H}} |\nabla f|^2\, dx\, dy}{\int\limits_{\Gamma\backslash\mathbb{H}} |f|^2 \dfrac{dx\,dy}{y^2}} \tag{4.1}$$

where the inf is taken over all non-zero complex valued $f \in L^2(\Gamma\backslash\mathbb{H})$ satisfying $\int_{\Gamma\backslash\mathbb{H}} f(dx\,dy)/y^2 = 0$. Here $\lambda_1(\Gamma)$ is the second smallest (0 is the smallest) eigenvalue of the Laplacian (for the hyperbolic metric) on Γ invariant functions on $\mathbb{H}$.

To construct surfaces S with large $m(P_S)$, we use a quaternion algebra. Let a, b be positive integers and let A be the quaternion algebra over $\mathbb{Q}$ generated by

$$1, i, j, k \quad \text{where } i^2 = a, j^2 = b, ij = -ji = k .$$

Choose a and b so that the form

$$\begin{aligned} N(X) &= Nm(X_0 + X_1 i + X_2 j + X_3 k) \\ &= X_0^2 - aX_1^2 - bX_2^2 + abX_3^2 \end{aligned}$$

does not represent zero for $(X_0, X_1, X_2, X_3) \in \mathbb{Q}^4$, $X \neq 0$. In this case A is a division algebra. Let $\tilde{\Gamma}$ be the group of all $X \in A(\mathbb{Z})$ with $N(X) = 1$. For p an odd prime define the congruence subgroup $\tilde{\Gamma}(p)$ of Γ by

$$\tilde{\Gamma}(p) = \{X \in \tilde{\Gamma} \mid X \equiv 1(p)\} .$$

We have an isomorphism of $\tilde{\Gamma}$ into $\mathrm{SL}_2(\mathbb{R})$ given by

$$X = X_0 + X_1 i + X_2 j + X_3 k \to \begin{bmatrix} X_0 + X_1\sqrt{a} & X_2 + X_3\sqrt{a} \\ b(X_2 - X_3\sqrt{a}) & X_0 - X_1\sqrt{a} \end{bmatrix} . \tag{4.2}$$

The corresponding lattices in $\mathrm{SL}_2(\mathbb{R})$ will be denoted by Γ and $\Gamma(p)$.

Since A is a division algebra, it is known that the quotient $\Gamma\backslash\mathbb{H}$ is compact, see [G-G-PS]. Also, $\Gamma(p)$ is torsion free (see below) so that $S_p = \Gamma(p)\backslash\mathbb{H}$ is a compact

Riemann surface. Its genus g_p is easily computed to be of the form

$$g_p = p(p-1)(p+1)\nu + 1 \tag{4.3}$$

where $\nu > 0$ depends on a and b, which we fix.

Next we estimate the girth of S_p, that is, the length of its shortest closed geodesic w.r.t. the hyperbolic metric. If

$$\alpha = X_0 + X_1 i + X_2 j + X_3 k \in \tilde{\Gamma}(p),$$

then

$$p \mid X_j \quad \text{for } j = 1, 2, 3 .$$

Also,

$$1 = N(\alpha) = X_0^2 - aX_1^2 - bX_2^2 + abX_3^2 .$$

Hence

$$X_0^2 \equiv 1 (p^2)$$

and therefore

$$X_0 \equiv \pm (p^2) . \tag{4.4}$$

If $\alpha \neq \pm 1$ then $X_0 \neq \pm 1$ (since Γ has no parabolic elements) and so if $\alpha \neq \pm 1$

$$|X_0| \geqq p^2 - 1 . \tag{4.5}$$

That is to say, if $\gamma \in \Gamma(p)$, $\gamma \neq 1$ then

$$|\mathrm{Trace}(\gamma)| \geqq 2p^2 - 2 . \tag{4.6}$$

In particular, all $\gamma \in \Gamma(p)$ are hyperbolic and so $S_p = \Gamma(p)\backslash \mathbb{H}$ is a compact Riemann surface. It follows that the girth of S_p satisfies

$$\mathrm{girth}(S_p) \geqq 2 \log p^2 . \tag{4.7}$$

Finally, concerning $\lambda_1(\Gamma(p))$, it is shown in [S-X] using elementary means that there is $\varepsilon_0 > 0$ (depending on a and b) such that

$$\lambda_1(\Gamma(p)) \geqq \varepsilon_0 \quad \text{for all } p . \tag{4.8}$$

Actually, using more sophisticated means which we appeal to in the next section, one can replace ε_0 by 3/16 in (4.8). For this section we do not need that.

To summarize, S_p satisfies

(i) $\mathrm{girth}(S_p) \geqq \frac{4}{3} \log(\mathrm{genus}(S_p)) - C$ (where C is independent of p),
(ii) $\mathrm{genus}(S_p) = \nu(p-1)(p+1)p + 1 = g_p$
(iii) $\lambda_1(S_p) \geqq \varepsilon_0 > 0$.

We remark that (i) should be compared with girth bounds for regular graphs – see [Sa].

Next we show that (i), (ii), (iii) above imply that

$$m(P_{S_p}) \geqq c \log g_p \tag{4.9}$$

for a suitable positive constant c.

To prove this we use the characterization (3.1). Let w be any closed 1-form on S_p with integral, not all zero, periods. Let

$$G(z) = e^{2\pi i \int_{z_0}^{z} w} \tag{4.10}$$

where z_0 is any fixed point of S_p. Observe that G is well defined as a function on S_p. Furthermore

$$\int_{S_p} |G|^2\, dV = 4\pi(g-1) =: V \tag{4.11}$$

$$\int_{S_p} |\nabla G|^2\, dV = 4\pi^2 \int_S w \wedge * w\,. \tag{4.12}$$

For $t \in [0, 2\pi)$ let

$$L_t = \{z \mid G(z) = e^{it}\}\,. \tag{4.13}$$

Computing everything from now on w.r.t. the hyperbolic metric, we have by the 'co-area' formula [Ch],

$$\int_{\{z \mid G(z) \in \alpha\}} |\nabla G|\, dV = \int_{3\pi/4}^{5\pi/4} \ell(L_t)\, dt \tag{4.14}$$

where $\ell(L_t)$ is the length of the level set L_t and $\alpha = \{e^{i\theta} \mid \frac{3\pi}{4} \leqq \theta \leqq \frac{5\pi}{4}\}$. For any t the connected components of L_t cannot all be homotopically trivial, for otherwise G would have zero winding number along all closed loops. Hence, according to (i), it follows that for each t

$$\ell(L_t) \geqq B \log g_p \tag{4.15}$$

for a constant $B > 0$.

This, together with (4.14), yields

$$\begin{aligned} \log g_p &\leqq B' \int_{\{z \mid G(z) \in \alpha\}} |\nabla G|\, dV \\ &\leqq B' A^{1/2} \left(\int_{\{z \mid G(z) \in \alpha\}} |\nabla G|^2\, dV \right)^{1/2} \\ &\leqq B'' A^{1/2} \|w\| \end{aligned} \tag{4.16}$$

where B' and B'' are constants,

$$A = \mathrm{Vol}(\{z \mid G(z) \in \alpha\})$$

and

$$\|w\|^2 = \int_{S_p} w \wedge * w\,.$$

By replacing $G(z)$ by $e^{i\xi} G(z)$ for suitable ξ, we can assume w.l.o.g. that

$$G_0 =: \int_{S_p} G(z)\, dV(z)$$

is real and nonnegative.

By the definition of A

$$0 \leqq G_0 = |G_0| \leqq V - A\,. \tag{4.17}$$

Write

$$G = \frac{1}{V} G_0 + G_1 \tag{4.18}$$

so that

$$\int_{S_p} G_1(z)\, dV(z) = 0 .$$

Using (iii), we conclude that

$$\int_{S_p} |\nabla G|^2\, dV = \int_{S_p} |\nabla G_1|^2\, dV \geqq \varepsilon_0 \int_{S_p} |G_1|^2\, dV .$$

Hence, from (4.17)

$$4\pi^2 \|w\|^2 = \int_{S_p} |\nabla G|^2\, dV \geqq \varepsilon_0 \left\{ \int_{S_p} |G|^2\, dV - \frac{1}{V}\left(\int_{S_p} G\, dV \right)^2 \right\}$$

$$\geqq \varepsilon_0 \left\{ V - \frac{1}{V}(V - A)^2 \right\}$$

and (4.16),

$$\|w\|^2 \geqq \varepsilon_0 A \geqq \varepsilon_0'(\log g_p)^2/\|w\|^2$$

for some ε_0' which is independent of p. Thus

$$\|w\|^2 \geqq \sqrt{\varepsilon_0'} \log g_p . \tag{4.19}$$

This proves (1.13) for g of the form $vp(p+1)(p-1)+1$.

To prove (1.13) in general, we could seek to vary a and b in the definition of the quaternion algebra and to vary the congruence groups, but this seems very complicated. Rather, we appeal to some results from additive analytic number theory. By use of the "circle method", see, for example Vaughan [Va], one can find large integers K and N_0 such that every $n \geqq N_0$ can be written as

$$n = g_1 + g_2 + \cdots + g_k \tag{4.20}$$

for some $k \leqq K$, g_j being of the form $vp(p+1)(p-1)+1$ and $g_j \geqq n^{1/2k}$. For each such g_j let S_j be the Riemann surface constructed as above. For t a complex parameter t in a neighborhood of 0 we can construct a surface S_t of genus $g = n = g_1 + \cdots + g_k$ as in [Fay, Chap. III], that is by pinching along cycles homologous to zero. See Fig. 11.

As shown in Fay [Fay] if we choose an appropriate canonical basis for the homology on S_t then as $t \to 0$, P_{S_t} will converge to the block matrix

$$\begin{pmatrix} P_{S_1} & & & \bigcirc \\ & P_{S_2} & & \\ & & \ddots & \\ \bigcirc & & & P_{S_k} \end{pmatrix}.$$

It follows that

$$\lim_{t\to 0} m(P_{S_t}) = \min_j\, (m(P_j)) \geqq c \log g_j \geqq c' \log n$$

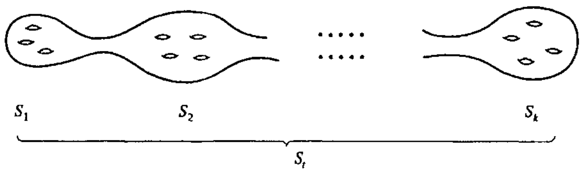

Fig. 11

with c' independent of n. Thus for t small enough we have a Riemann surface S_t satisfying the lower bound in (1.13).

5

To prove (1.17) we will use the following fact about congruence lattices, which follows from the theory of automorphic forms, see [Vi], for example. If $\Gamma \leqq \mathrm{SL}_2(\mathbb{R})$ is a congruence lattice, then

$$\lambda_1(\Gamma) \geqq 3/16 . \tag{5.1}$$

It is known [R1, Se] that, given $\varepsilon > 0$ and Γ any lattice in $\mathrm{SL}_2(\mathbb{R})$, there is a $\Gamma' < \Gamma$ of finite index such that

$$\lambda_1(\Gamma') < \varepsilon . \tag{5.2}$$

Our purpose is to prove the following sharp quantitative version of (5.2) which, together with (5.1), implies (1.17).

Given $\varepsilon > 0$, then for all but a finite number of arithmetic lattices, every arithmetic lattice has a subgroup Γ^* of index at most 2 for which

$$\lambda_1(\Gamma^*) < \varepsilon . \tag{5.3}$$

We first prove the version (1.18). Let $g(\Gamma)$ denote the genus of $\Gamma \backslash \mathbb{H}$. We recall that Zograf [Z] has shown that for any Γ

$$\lambda_1(\Gamma) < \frac{8\pi(g(\Gamma)+1)}{\mathrm{Area}(\Gamma \backslash \mathbb{H})} . \tag{5.4}$$

Thus, for the purpose of proving (5.3), we may assume that

$$\frac{8\pi(g(\Gamma)+1)}{\varepsilon} \geqq \mathrm{Area}(\Gamma \backslash \mathbb{H}) . \tag{5.5}$$

Our aim is to bound $\mathrm{Area}(\Gamma \backslash \mathbb{H})$ for Γ's for which (5.3) fails. To this end, from (5.5) we can assume $g(\Gamma) \geqq 2$. We now define a test form on $\Gamma \backslash \mathbb{H}$. For this we temporarily ignore the hyperbolic metric and regard $\Gamma \backslash \mathbb{H}$ as a surface with complex structure. Then $\Gamma \backslash \mathbb{H}$ after compactification and choice of uniformizers at the elliptic fixed points is a Riemann surface $\tilde{S}$ of genus g. It follows from (1.13) that $\tilde{S}$ carries a closed (harmonic) 1-form w, with integral periods, and at least one period equal one, for which

$$\int_{\tilde{S}} w \wedge * w \leqq \frac{3}{\pi} \log(4g+3). \tag{5.6}$$

w lifts to $\mathbb{H}$ giving a Γ invariant 1-form and $\int_{\tilde{S}} w \wedge * w = \int_S w \wedge * w$. For $0 \leqq \theta \leqq 1$ and $z \in \mathbb{H}$ set

$$G(z) = e^{2\pi i \theta \int_{z_0}^{z} w}.$$

Then

$$\frac{G(\gamma z)}{G(z)} = \chi_\theta(\gamma), \tag{5.7}$$

where

$$\chi_\theta \in \mathrm{Hom}(\Gamma, S^1).$$

In fact $\theta \to \chi_\theta$ is a (nontrivial) homomorphism of S^1 into the group of characters of Γ. $\chi_{1/2}$ is of order 2 and if $\tilde{\Gamma} = \ker \chi_{1/2}$, then clearly $\tilde{\Gamma}$ is an index 2 subgroup of Γ. The spectrum of $\tilde{\Gamma} \backslash \mathbb{H}$ consists of the union of the spectra of $\Gamma \backslash \mathbb{H}$ and of $L^2(\Gamma \backslash \mathbb{H}, \chi_{1/2})$; that is, functions on $\mathbb{H}$ satisfying $f(\gamma z) = \chi_{1/2}(\gamma) f(z)$. Thus (using again the hyperbolic metric)

$$\lambda_1(\tilde{\Gamma}) \leqq \inf \frac{\int_{\Gamma \backslash \mathbb{H}} |\nabla f|^2 \, dV}{\int_{\Gamma \backslash \mathbb{H}} |f|^2 \, dV}$$

where the inf is over all non-zero f satisfying $f(\gamma z) = \chi_{1/2}(\gamma) f(z)$. Now the function

$$G(z) = e^{i\pi \int_{z_0}^{z} w}$$

is a perfectly good test function for the latter. Moreover,

$$\int_{\Gamma \backslash \mathbb{H}} |G|^2 \, dV = \mathrm{Area}(\Gamma \backslash \mathbb{H}) \geqq 4\pi(g-1).$$

Moreover, by (5.6)

$$\int_{\Gamma \backslash \mathbb{H}} |\nabla G|^2 \, dV = \pi^2 \|w\|^2 \leqq \pi^2 \frac{3}{\pi} \log(4g+3).$$

Thus

$$\lambda_1(\tilde{\Gamma}) \leqq \frac{3 \log(4g+3)}{4(g-1)}. \tag{5.8}$$

Hence, if g is large enough, say greater than $C = C(\varepsilon)$, we have

$$\lambda_1(\tilde{\Gamma}) < \varepsilon. \tag{5.9}$$

We conclude that (5.3) is valid if $g(\Gamma) \geqq C(\varepsilon)$. This proves (1.18). Furthermore it shows that the arithmetic Γ's for which (5.3) fails must, according to (5.5), satisfy

$$\mathrm{Area}(\Gamma \backslash \mathbb{H}) \leqq \frac{8\pi(C(\varepsilon)+1)}{\varepsilon}. \tag{5.10}$$

By a result of Borel [Bo], the set of arithmetic lattices satisfying (5.10) is finite. This completes the proof of (5.3).

Appendix 0: Polarizations and quadratic forms

A polarization for a lattice $\Lambda \subset \mathbb{C}^g$ is a positive definite Hermitian form $H(\xi, \eta)$ for which the alternating form $E = \operatorname{Im} H$ is integral on $\Lambda \times \Lambda$. In this case $(\Lambda \backslash \mathbb{C}^g, H)$ is a polarized abelian variety. If, moreover, $pfaf(E) = 1$ we say the polarization (or the abelian variety) is principal. In this case we can choose a $\mathbb{Z}$ basis for Λ such that E is of the form $J = \left[\begin{smallmatrix} 0 & I \\ -I & 0 \end{smallmatrix}\right]$. Λ and H can be brought to standard form [M], that is

$$\left.\begin{aligned} &\Lambda = m + Zn, \quad m, n \in \mathbb{Z}^g \\ &Z = X + iY \in \mathbb{H}^g \\ &H(\xi, \eta) = {}^t\bar{\xi} Y^{-1} \eta \end{aligned}\right\}. \tag{A.0.1}$$

Relative to the basis $[I, Z]$ of $\mathbb{C}^g = \mathbb{R}^{2g}$ $E((m, n), (m', n')) = {}^tmn' - {}^tnm'$. That is $E = J$. In this form Z and Z' correspond to the same principally polarized abelian variety iff

$$Z' = (AZ + B)(CZ + D)^{-1}$$

where

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \Gamma = \mathrm{Sp}(2g, \mathbb{Z}) .$$

Associated to Λ and H as above we have a positive definite (real) quadratic form F on $\mathbb{C}^g = \mathbb{R}^{2g}$ given by

$$F(\zeta) = H(\zeta, \zeta), \quad \zeta \in \mathbb{C}^g . \tag{A.0.2}$$

If we choose as above, the basis for Λ so that $E = J$ then F has a matrix representation P which is a symplectic quadratic form. Indeed if $\Lambda = m + Zn$, $m, n \in \mathbb{Z}^g$ and H is as above, then

$$\begin{aligned} F((m, n) &= {}^t\overline{(m + Zn)}\, Y^{-1}(m + Zn) \\ &= {}^t(m + Xn)\, Y^{-1}(m + Xn) + {}^tnYn . \end{aligned}$$

That is the matrix P of F is given by

$$P_Z = \begin{bmatrix} I & 0 \\ X & I \end{bmatrix} \begin{bmatrix} Y^{-1} & 0 \\ 0 & Y \end{bmatrix} \begin{bmatrix} I & X \\ 0 & I \end{bmatrix} .$$

This is the association $Z \leftrightarrow P_Z$ described in the Introduction. It is one-to-one and $Z \sim Z' (\operatorname{mod} \Gamma)$ iff $P_Z \sim P_{Z'} (\operatorname{mod} \Gamma)$.

In terms of the polarization $A = (\Lambda, H)$ note that the key invariant m in (1.6) is defined in a basis free way by

$$m(A) = \min_{0 \neq \lambda \in \Lambda} H(\lambda, \lambda) . \tag{A.0.3}$$

Finally let S be a compact Riemann surface and $\operatorname{Jac}(S)$ its principally polarized Jacobian in the form $(\mathbb{Z}^g + \Pi \mathbb{Z}^g) \backslash \mathbb{C}^g$ where $\Pi \in \mathbb{H}^g$ is the period matrix of S.

Write $\Pi = X + iY$ then according to the formulas on p. 58–61 in [F-K] we have that the symplectic quadratic form P_S, as defined in (1.5), is given by

$$P_S = \begin{bmatrix} Y + XY^{-1}X & -XY^{-1} \\ -Y^{-1}X & Y^{-1} \end{bmatrix} \tag{A.0.4}$$

so that

$$P_S = P_\Pi^{-1} = {}^tJP_\Pi J\,. \tag{A.0.5}$$

Since $J \in \Gamma$, P_S and P_Π are equivalent symplectic quadratic forms and our various definitions are consistent.

Appendix 1: Some explicit examples

For $g = 1$ it is well known that $\gamma_2 = 2/\sqrt{3}$ corresponding to the form with matrix $P = 2/\sqrt{3}\left[\begin{smallmatrix} 1 & 1/2 \\ 1/2 & 1 \end{smallmatrix}\right]$. P is symplectic (this is automatic in dimension 2) and since $\mathcal{M}_1 = \mathcal{A}_1$

$$\eta_2 = \delta_2 = \gamma_2 = \frac{2}{\sqrt{3}}\,. \tag{A.1.1}$$

Now $Z_P = e^{i\pi/3}$ and so $m(P_S)$ is extremized by the curve $L\backslash\mathbb{C}$ where $L = \{m + e^{i\pi/3}n \mid m, n \in \mathbb{Z}\}$.

The situation for $g = 2$ is similar. The extremal quadratic form in $\mathbb{R}^4$ is the Gram matrix P_{D_4} of the lattice D_4 (see [C-S]). In Appendix 2 it is shown that P_{D_4} is symplectic and hence $\delta_4 = \gamma_4 = \sqrt{2}$. Since $\overline{t(\mathcal{M}_2)} = \mathcal{A}_2$ we have

$$\eta_4 = \gamma_4 = \delta_4 = \sqrt{2}\,. \tag{A.1.2}$$

In fact $Z_{P_{D_4}} \in \mathbb{H}^2$ corresponds to the Jacobian of the curve $y^2 = x^5 - x$ (see Gross [Gr] where a discussion is given of how this and other lattices arise from group representations, see also Bolza [Bol]).

The extremal quadratic form Q on $\mathbb{R}^6$ is the Gram matrix of the lattice E_6 [C-S] and

$$m(Q) = \gamma_6 = \left(\frac{64}{3}\right)^{1/6} = 1.665\ldots. \tag{A.1.3}$$

Moreover E_6 is not self dual. Since any symplectic quadratic form P is self dual as it satisfies

$$P^{-1} = {}^tJPJ\,,$$

it follows that Q is not symplectic. This together with $\overline{t(\mathcal{M}_3)} = \mathcal{A}_3$, implies that

$$\eta_6 = \delta_6 < \gamma_6\,.$$

A natural candidate for η_6 comes from the Jacobian of the Klein curve K: $xy^3 + yz^3 + zx^3 = 0$. Its Jacobian has been studied by Serre (see Mazur [Ma]) and its polarized Jacobian has its symplectic form corresponding to $A_6^{(2)}$ in Appendix 2.

Moreover

$$m(K) = \frac{4}{\sqrt{7}}\,.$$

Hence

$$1.511\ldots = \frac{4}{\sqrt{7}} \leqq \eta_6 = \delta_6 < \gamma_6 = 1.665\ldots$$

with equality on the left being quite likely.

For $g = 4$ the extremal quadratic form in $\mathbb{R}^8$ is P_{E_8} – the Gram matrix of the lattice E_8[C-S]. P_{E_8} is symplectic, see Appendix 2 where $Z_{E_8} \in \mathbb{H}^4$ is given. Hence $\delta_8 = \gamma_8 = 2$. As was pointed out to us by F. Rodriguez Villegas, one can use Torelli's theorem in the form proved by Weil (see Mazur [Ma]) to deduce Z_{E_8} is not a Jacobian. Indeed, according to Appendix 2, the automorphism group of the polarized abelian variety corresponding to Z_{E_8} is of order 46080. If Z_{E_8} were a Jacobian of a curve S then according to Torelli's theorem the automorphism group of S would have order 46080 or 46080/2. However, according to Hurwitz's well known theorem, the automorphism group of a curve of genus 4 has order at most 252. Thus, since Z_{E_8} is not a Jacobian and is irreducible, we have

$$\eta_8 < 2 = \gamma_8 = \delta_8\,.$$

We are not aware of a good candidate for η_8. For comparison, we mention that the inequality (3.20) yields the following bounds.

$$\eta_4 \leqq 1.645\ldots, \quad \eta_6 \leqq 1.888\ldots, \quad \eta_8 \leqq 2.058\ldots.$$

Finally, it is shown in Appendix 2 that the Leech Lattice Λ_{24} is symplectic. Let P_Λ and Z_Λ be the corresponding Gram and Siegel matrices (see Appendix 2). Now according to (1.13)

$$\eta_{24} \leqq \frac{3}{\pi}\log 51 = 3.7546\ldots.$$

(With (3.20) we even get $\eta_{24} \leqq 2.696\ldots$.) Since $m(P_\Lambda) = 4$ it follows that Z_Λ is not a Jacobian. We conclude that $\eta_{24} < \delta_{24}$ and (since in all likelihood the Leech lattice is m-extremal) that probably $\delta_{24} = \gamma_{24} = 4$. Again, one can use Torelli's theorem to show that Z_Λ is not a Jacobian. Indeed, the automorphism group of the polarized abelian variety $L \backslash \mathbb{C}^{12}$, where $L = \{m + Z_\Lambda n \,|\, m, n \in \mathbb{Z}^{12}\}$, is of order 2012774400 (see Appendix 2). This number is substantially larger than the Hurwitz bound for genus 12.

Moving to larger genus, our condition (1.13) together with (1.12) is effective for $g \geqq 38$ (respectively, for $g \geqq 23$ if we use (3.20)). Given a random P in $\mathfrak{h}_{2g}$ our test of whether P is a Jacobian by computing $m(P)$ is, according to (1.15), robust. We point out, however, that the problem of suitably numerically estimating $m(P)$ for large g is a difficult one even with the recent developments of such fast algorithms as for example in [L-L-L].

We give some explicit families of abelian varieties which are not Jacobians using (1.13). If $Z = X + iY \in \mathbb{H}^g$ and P_Z is the corresponding symplectic form, then from

$$(r,s)P_Z^t(r,s) = (r+sX)\,Y^{-1}(r+sX)^t + sY^ts$$

we see that

$$m(Z) =: m(P_Z) \geqq \min(m(Y), m(Y^{-1})) . \tag{A.1.4}$$

Hence to find Z's with large m we look for self-dual forms $Y \in P_g$.

Recall the definition of the Barnes–Wall lattices [B-W]. Let $g = 2^n$, $n \geqq 4$, $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ the standard basis for $V = \{0, 1\}^n$ the vector space of dimension n over $\mathbb{Z}/2\mathbb{Z}$. Then $|V| = g$ so we may realize the integer lattice $\mathbb{Z}^g$ as $\{x = x(\alpha)\}$, x having coordinates $x(\alpha)$ where $\alpha \in V$. For $W \subset V$ define the lattice vector x_w by

$$x_w(\alpha) = \begin{cases} 1 & \text{if } \alpha \in W \\ 0 & \text{if } \alpha \notin W . \end{cases} \tag{A.1.5}$$

The g vectors

$$2^{[(n-r)/2]} x_{W_r} \quad 0 \leqq r \leqq n \tag{A.1.6}$$

where W_r ranges over all subgroups of V of dimension r having a subset of $\varepsilon_1, \ldots, \varepsilon_n$ as basis, form a basis for a lattice $\tilde{\Lambda}_g \subset \mathbb{Z}^g$. If we scale $\tilde{\Lambda}_g$ to have covolume 1 in $\mathbb{R}^g$ we obain the Barnes–Wall lattice Λ_g. The lattice Λ_g is self dual and its Gram matrix Q_g satisfies

$$m(Q_g) = \left(\frac{g}{2}\right)^{1/2} = m(Q_g^{-1}) . \tag{A.1.7}$$

Consider $Z = X + iY \in \mathbb{H}^g$ satisfying for instance

$$\tfrac{3}{4} Q_g \leqq Y \leqq \tfrac{4}{3} Q_g . \tag{A.1.8}$$

Here $P \leqq R$ means ${}^t\xi P \xi \leqq {}^t\xi R \xi$ for all $\xi \in \mathbb{R}^g$. According to (A.1.7) and (A.1.4) we conclude that for $g = 2^n$, $n \geqq 7$ and Z satisfying (A.1.8)

$$m(Z) > \frac{3}{\pi} \log(4g + 3) . \tag{A.1.9}$$

Hence from (1.3) we conclude:

If $g = 2^n$ *with* $n \geqq 7$, *and if* $Z \in \mathbb{H}^g$ *satisfies* (A.1.8) *then* Z *is not a period matrix.* (A.1.10)

Appendix 2: D_4, E_8, Leech and certain other lattices are symplectic

J.H. Conway and N.J.A. Sloane

A $2n$-dimensional lattice Λ is symplectic if and only if it has a Gram matrix P of the form

$$P = \begin{bmatrix} I & 0 \\ -X & I \end{bmatrix} \begin{bmatrix} Y & 0 \\ 0 & Y^{-1} \end{bmatrix} \begin{bmatrix} I & -X \\ 0 & I \end{bmatrix},$$

with X, Y symmetric, $Y > 0$ (this is equivalent to the version used in the introduction and the translation to $Z \in \mathbb{H}^g$ can easily be achieved using the version following (1.3)). Any two-dimensional lattice may be scaled to be symplectic. We show that suitably scaled versions of the lattices (named as in [C-S]) $D_4, D_{4m}^+ \, (m \geqq 1), A_6^{(2)}, E_8, K_{12}, \Lambda_{16}, \Lambda_{24}$ (the Leech lattice) and certain others, are symplectic.

Theorem. *A lattice Λ is symplectic if and only if there is an orthogonal transformation $\imath$ on the space such that $\imath^2 = -1$ and $\imath(\Lambda)$ is the dual lattice Λ^*.*

Proof. The following argument is rather more complicated than necessary but also yields a basis. (Only if:) We regard the matrix $\begin{bmatrix} Y & 0 \\ 0 & Y^{-1} \end{bmatrix}$ as the Gram matrix of the direct sum of a lattice L in a space V and a copy $\bar{L}^*$ of the dual lattice L^* in a space $\bar{V}$. Let $e_1, \ldots, e_n$ be an integral basis for L, $e_1^*, \ldots, e_n^*$ the dual basis for L^*, and $\bar{e}_1, \ldots, \bar{e}_n$ and $\bar{e}_1^*, \ldots, \bar{e}_n^*$ the corresponding bases for $\bar{L}$ and $\bar{L}^*$. The matrix $\begin{bmatrix} I & 0 \\ -X & I \end{bmatrix}$ represents a base change under which $\bar{e}_i^*$ is replaced by

$$- \sum_{j=1}^{n} x_{ij} e_j + \bar{e}_i^* \quad (i = 1, \ldots, n) .$$

Then $\begin{bmatrix} I & 0 \\ -X & I \end{bmatrix} \begin{bmatrix} Y & 0 \\ 0 & Y^{-1} \end{bmatrix} \begin{bmatrix} I & -{}^tX \\ 0 & I \end{bmatrix}$ is the Gram matrix of a lattice whose basis vectors are

$$e_i, \quad - \sum_{j=1}^{n} x_{ij} e_j + \bar{e}_i^* \quad (i = 1, \ldots, n) . \tag{A.2.1}$$

The dual basis is

$$e_i^* + \sum_{j=1}^{n} x_{ji} \bar{e}_j, \quad \bar{e}_i \quad (i = 1, \ldots, n) . \tag{A.2.2}$$

The symmetry condition $x_{ij} = x_{ji}$ (all i, j) holds precisely when the transformation

$$\imath = \sum_{i=1}^{n} (a_i e_i + b_i \bar{e}_i) \mapsto \sum_{i=1}^{n} (b_i e_i - a_i \bar{e}_i)$$

maps Λ to Λ^*. (If:) Let $\imath$ be as stated in the theorem. We first select vectors $e_1', \ldots, e_n'$ which are linear combinations of vectors in Λ, choosing e_k' $(k = 1, \ldots, n)$ to be orthogonal to $\imath(e_1'), \ldots, \imath(e_{k-1}')$. e_k' is automatically orthogonal to $\imath(e_k')$. Let $V = \mathbb{R}\langle e_1', \ldots, e_n' \rangle$, $\bar{V} = \mathbb{R}\langle \imath(e_1'), \ldots, \imath(e_n') \rangle$. Let $e_1, \ldots, e_n$ be a basis for $\Lambda \cap V$, let $\bar{e}_1, \ldots, \bar{e}_n$ be their $\imath$-images in $\bar{V}$, and let $\{e_i^*\} \subseteq V$, $\{\bar{e}_i^*\} \subseteq \bar{V}$ be the dual bases. Any $u \in \Lambda$ can be written $u = v + w$, $v \in V$, $w \in \bar{V}$. Since $\imath$ maps Λ to Λ^*, there are u's whose projections w include all the $\bar{e}_i^*$. Thus, we can find a basis for Λ of the form (A.2.1), whose dual basis is (A.2.2). But applying $\imath$ to (A.2.1) we obtain a second basis for Λ^*, namely

$$- e_i^* + \sum_{j=1}^{n} x_{ij} \bar{e}_j, \quad \bar{e}_i \quad (i = 1, \ldots, n) . \tag{A.2.3}$$

Comparing (A.2.2), (A.2.3), we see that x_{ij} and x_{ji} differ by integers, and so, by adding multiples of the $\bar{e}_j$'s, X may be taken to be symmetric.

Corollary. *Let Λ be a classically integral lattice. Then Λ is symplectic if and only if it is unimodular and has an automorphism $\imath$ satisfying $\imath^2 = -1$.*

The theorem shows that D_{4m}^{+} $(m \geqq 1)$, $A_6^{(2)}$, E_8 and Λ_{24} are symplectic (an ι for Λ_{24} is given in [C-S, p. 178]). Some explicit matrices X, Y are as follows:

D_4: $$X = \frac{1}{3}\begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix}, \qquad Y = \frac{1}{\sqrt{2}}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

$A_6^{(2)}$: $$-X = \frac{1}{12}\begin{bmatrix} 4 & -2 & 4 \\ -2 & -1 & 6 \\ 4 & 6 & 8 \end{bmatrix}, \qquad Y = \frac{1}{\sqrt{7}}\begin{bmatrix} 4 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 4 \end{bmatrix},$$

E_8: $$-X = \frac{1}{2}\begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & 1 \\ -1 & 0 & 1 & 2 \end{bmatrix}, \qquad Y = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & -1 \\ 1 & 0 & -1 & 2 \end{bmatrix},$$

Λ_{24}: $$-X = \frac{1}{2}\begin{bmatrix} 14 & 3j_{11} \\ 3^t j_{11} & C \end{bmatrix}, \qquad Y = \begin{bmatrix} 24 & -10 j_{11} \\ -10^t j_{11} & 2I_{11} + 4J_{11} \end{bmatrix},$$

where j_{11} is a vector of 11 1's, J_{11} is an 11×11 matrix of 1's, and C is a retro-circulant with first row $(1, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0)$. An integral basis for Λ_{24} corresponding to this pair X, Y consists of the rows of the matrix

$$\begin{bmatrix} -4 & 4j_{11} & 0 & 0 \\ 2^t j_{11} & 4I_{11} - 2J_{11} & 0 & 0 \\ 5 & j_{11} & 9 & j_{11} \\ 0 & 2C & 2^t j_{11} & 2I_{11} \end{bmatrix}.$$

For E_8 the subgroup of the automorphism group that commutes with ι has structure $4.2^4 S_6$ and order 46080. For Λ_{24} the corresponding group is $(4 \times G_2(4))$: 2 of order 2012774400.

Acknowledgements. We would like to thank W. Duke, N. Elkies, R. Gunning, R. Pink, F. Rodriguez-Villegas, and B. Randol for discussions related to this work. We also thank the referee for the comments.

References

[B-W] Barnes, E.S., Wall, G.E.: Some extreme forms defined in terms of abelian groups. J. Aust. Math. Soc. **1**, 47–63 (1959)

[B-M-S] Bass, H., Milnor, J., Serre, J.P.: Solutions of the congruence subgroup problem for SL(n) ($n \geqq 3$) and Sp($2n$) ($n \geqq 2$). Publ. Math., Inst. Hautes Étud. Sci. **33**, 59–137 (1967)

[Be] Beardon, A.F.: The Geometry of discrete groups. (Grad. Texts Math., vol. 91) Berlin Heidelberg New York: Springer 1983

[Bol] Bolza, O.: On binary sextics with linear transformations into themselves. Am. J. Math. **10**, 47–70 (1888)

[Bo] Borel, A.: Commensurability classes and volumes of hyperbolic 3-manifolds. Ann. Sc. Norm. Super. Pisa, IV. Ser. **8**, 1–33 (1981)

[Bu] Buser, P.: Geometry and spectra of compact Riemann surfaces. Boston Basel Stuttgart: Birkhäuser 1992
[Ch] Cheeger, J.: A lower bound for the smallest eigenvalue of the Laplacian. In: Gunning, R. (ed.) Problems in analysis, pp. 195–199. Princeton: Princeton University Press 1970
[C-S] Conway, J.H., Sloane, N.J.A.: Sphere packings, lattices and groups. (Grundlehren Math. Wiss., vol. 290) Berlin Heidelberg New York, 1988
[E] Epstein, D.B.A.: Curves on 2-manifolds and isotopies. Acta Math. **115**, 83–107 (1966)
[F] Farkas, H.: Schottky-Jung theory. In: Ehrenpreis, L., Gunning, R.C. (eds.) Theta functions, Bowdoin 1987. (Proc. Symp. Pure Math., vol. 49, 459–483) Providence, RI: Am. Math. Soc. 1989
[F-K] Farkas, H., Kra, I.: Riemann surfaces. Berlin Heidelberg New York: Springer 1980
[F-L-P] Fathi, A., Laudenbach, F., Poénaru, V.: Travaux de Thurston sur les surfaces. Seminaire Orsay. Astérisque **66–67** (1979)
[Fay] Fay, J.D.: Theta functions on Riemann surfaces. (Lect. Notes Math., vol. 352) Berlin Heidelberg New York: Springer 1973
[G-G-PS] Gel'fand, I.M., Graev, M.I., Pyatetskii-Shapiro, I.I.: Representation theory and automorphic functions. (Generalized functions, vol. 6) Philadelphia London Toronto: Saunders 1969
[Gro] Gromov, M.: Systoles and intersystolic inequalities. (Preprint 1993)
[Gr] Gross, B.H.: Group representations and lattices. J. Am. Math. Soc. **3**, no. 4, 929–960 (1990)
[Gru-Le] Gruber, P.M., Lekkerkerker, C.G.: Geometry of numbers, Amsterdam: North Holland 1987
[Gu] Gunning, R.C.: Some curves in abelian varieties. Invent. Math. **66**, 377–389 (1982)
[I] Igusa, J.: On the irreducibility of Schottky's divisor. J. Fac. Sci. Tokyo **28**, 531–544 (1982)
[K-L] Kabatiansky, G., Levenshtein, V.: Bounds for packings of a sphere in space. Prob. Peradachi Inf. **14**, no. 1 (1978)
[L] Lang, S.: Introduction to algebraic and abelian functions. (Grad. Texts Math., vol. 89) Berlin Heidelberg New York: Springer 1982
[L-L-L] Lenstra, A.K., Lenstra, H.W. Jr., Lovász, L.: Factoring polynomials with rational coefficients. Math. Ann. **261**, 515–534 (1982)
[Ma] Mazur, B.: Arithmetic on curves. Bull. Am. Math. Soc. **14**, 207–259 (1986)
[M] Mumford, D.: Curves and their Jacobians. Ann Arbor. University of Michigan Press 1976
[Ra] Raghunathan, M.S.: On the congruence subgroup problem. Publ. Math., Inst. Hautes Étud. Sci. **46**, 107–161 (1976)
[R1] Randol, B.: Small eigenvalues of the Laplace operator on compact Riemann surfaces. Bull Am. Math. Soc. **80**, 996–1000 (1974)
[R2] Randol, B.: Cylinders in Riemann surfaces. Coment. Math. Helv. **54**, 1–5 (1979)
[Rn] Rankin, R.A.: The modular group and its subgroups. Madras: Ramanujan Inst. 1969
[Sa] Sarnak, P.: Some applications of modular forms. (Cambridge Tracts Math., vol. 99) Cambridge: Cambridge University Press 1990
[S-X] Sarnak, P., Xue, X.: Bounds for multiplicities of automorphic representations. Duke Math. J. 64 (no. 1), 207–227 (1991)
[Se] Selberg, A.: On the estimation of Fourier coefficients of modular forms. In: Collected Works, pp. 506–520. Berlin Heidelberg New York: Springer 1988
[Sh] Shiota, T.: Characterization of Jacobian varieties in terms of soliton equations. Invent. Math. **83**, 333–382 (1986)
[Si] Siegel, C.L.: Symplectic geometry, Amer. J. Math. **65**, 1–86 (1943)
[T] Thurston, W.: The geometry and topology of 3-manifolds. (Princeton University Notes
[Va] Vaughan, R.C.: The Hardy Littlewood method. (Cambridge Tracts Math., vol. 80) Cambridge: Cambridge University Press 1981
[Vi] Vignéras, M.F.: Quelques remarques sur la conjecture $\lambda_1 \geqq 1/4$. In: Bertin, M.-J. (ed.) Sém. de théorie des nombres, Paris 1981–1982. (Prog. Math., vol. 38, pp. 321–343) Boston Basel Stuttgart: Birkhäuser 1983
[Z] Zograf, P.: A spectral proof of Rademacher's conjecture for congruence subgroups of the modular group. J. Reine Angew. Math. **414**, 113–116 (1991)