Pacific Journal of Mathematics

Volume 202 No. 1 January 2002

PACIFIC JOURNAL OF MATHEMATICS

http://www.pjmath.org

Founded in 1951 by

E. F. Beckenbach (1906–1982) F. Wolf (1904–1989)

See inside back cover or www.pjmath.org for submission instructions.

Regular subscription rate for 2006: $425.00 a year (10 issues). Special rate: $212.50 a year to individual members of supporting institutions. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. Prior back issues are obtainable from Periodicals Service Company, 11 Main Street, Germantown, NY 12526-5635. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and the Science Citation Index.

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 969 Evans Hall, Berkeley, CA 94720-3840 is published monthly except July and August. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

PUBLISHED BY PACIFIC JOURNAL OF MATHEMATICS
at the University of California, Berkeley 94720-3840
A NON-PROFIT CORPORATION
Typeset in LaTeX

PACIFIC JOURNAL OF MATHEMATICS
Vol. 202, No. 1, 2002

SPECTRUM AND ASYMPTOTICS OF THE BLACK–SCHOLES PARTIAL DIFFERENTIAL EQUATION IN (L^1, L^∞)-INTERPOLATION SPACES

Wolfgang Arendt and Ben de Pagter

Let E be an (L^1, L^∞)-interpolation space. Then $(T_E(t)f)(x) = f(e^{-t}x)$ defines a group on E. It is strongly continuous if and only if E has order continuous norm. In any case, a generator A_E can be associated with T_E. It is shown that its spectrum is the strip $\{\underline{\alpha}_E \leq Re\,\lambda \leq \overline{\alpha}_E\}$, where $\underline{\alpha}_E$ and $\overline{\alpha}_E$ are the Boyd indices of E. The operator $B_E = (A_E)^2$ generates a holomorphic semigroup which governs the Black–Scholes partial differential equation $u_t = x^2 u_{xx} + x u_x$, whose well-posedness, spectrum and asymptotics in E are studied.

0. Introduction.

Let E be an (L^1, L^∞)-interpolation space on $(0, \infty)$, $\mathbb{R}$ or $\mathbb{T}$. Then the upper and lower Boyd indices $\underline{\alpha}_E$ and $\overline{\alpha}_E$ are of great importance. For example, the Hilbert transform is bounded on E if and only if $0 < \underline{\alpha}_E$ and $\overline{\alpha}_E < 1$. Also norm convergence of the Fourier series can be expressed in terms of the Boyd indices (see [**BS**]). In his paper [**Bo**], Boyd computes the spectrum of the Cesaro operator in terms of the Boyd indices. Here we consider a natural one-parameter group of dilations $(T_E(t))_{t\in\mathbb{R}}$ on E. It turns out that the Boyd indices are just the growth bounds (or exponential bounds) of this group.

To be more precise, we consider an (L^1, L^∞)-interpolation space E on $(0, \infty)$ througout this article. The group T_E on E is defined by

$$(T_E(t)f)(x) = f(e^{-t}x)$$

for all $f \in E,\ t \in \mathbb{R},\ x > 0$. Now the first problem is that T_E is not strongly continuous, in general. In fact, one of our main results says that T_E is a C_0-group if and only if E has order continuous norm.

Still it is possible to associate a generator A_E to T_E without any further condition on the space, and we show that its spectrum is the strip

$$\sigma(A_E) = \{\lambda \in \mathbb{C} : \underline{\alpha}_E \leq Re\,\lambda \leq \overline{\alpha}_E\}.$$

Thus the spectrum of A_E varies very much in function of the space E. It turns out that the Cesaro operator is just $(1 - A_E)^{-1}$. Thus it is bounded if and only if $\overline{\alpha}_E < 1$. In that case we obtain its spectrum just by applying the result above on the spectrum of A_E.

Of particular interest is the operator $B_E = (A_E)^2$. In fact, B_E is a degenerate elliptic operator given by $(B_E f)(x) = x^2 f''(x) + x f'(x)$ with suitable domain. As a consequence of the results on A_E we obtain much information on B_E. It always generates a generalized holomorphic semigroup V_E on E. So this semigroup gives the solution of the Black-Scholes partial differential equation

$$u_t = x^2 u_{xx} + x u_x. \tag{BS}$$

We show that the semigroup V_E is strongly continuous if and only if E has order continuous norm. Nevertheless, one of the main results says that T_E as well as V_E are always $\sigma(E, E_n')$-continuous, where E_n' is the Köthe dual of E; i.e., the space of all functionals given by a measurable function. This allows us to formulate precisely well-posedness for (BS) in E. Finally, we consider perturbations of the operator B_E. The results imply in particular well-posedness of the more general equation

$$u_t = \alpha x^2 u_{xx} + \beta x u_x + \gamma u$$

where $\alpha > 0$ is a constant and $\beta, \gamma \in L^\infty(0, \infty)$.

Because of its importance in mathematical finance (see [**BlSc**]), the Black-Scholes partial differential equation has been investigated most recently. We refer to Gozzi, Monte, Vespri [**GMV**], Barucci, Gozzi, Vespri [**BGV**] and Colombo, Giuli, Vespri [**CGV**] for further information. We would like to emphasize that the motivation for this work lies in the interesting relations between properties of interpolation spaces and the semigroups considered here. It is not at all a contribution to modelling in mathematical finance.

The paper is organized in the following way: After some preliminaries we show in Section 2 that the semigroup T_E is strongly continuous if and only if E has order continuous norm. In that case we can use a result of Greiner [**G**] to determine the spectrum of E. In the less conventional situation where E does not have order continuous norm we use the theory of resolvent positive operators and integrated semigroups. Now the situation is much more complicated, and Section 3 is devoted to the generalization of Greiner's decomposition theorem to resolvent bipositive operators. In Section 4 we prove the results on the spectrum in the general case. Here it is also shown that the semigroup T_E is $\sigma(E, E_n')$-continuous. In Section 5 we investigate the Black-Scholes operator $B_E = (A_E)^2$. Its perturbations are studied in Section 6.

1. Preliminaries.

On the interval $(0,\infty)$ we consider Lebesgue measure (dm or dx). For a Borel measurable function $f : (0,\infty) \to \mathbb{C}$ the **distribution function** is defined by $d_{|f|}(\lambda) = m\{t \in (0,\infty) : |f(t)| > \lambda\}$ for $\lambda > 0$. We will consider only functions f for which $d_{|f|}(\lambda) < \infty$ for some $\lambda > 0$. The space of all such functions will be denoted by $\mathcal{S}_0(0,\infty)$. For $f \in \mathcal{S}_0(0,\infty)$ we define

$$f^*(t) = \inf\{\lambda > 0 : d_{|f|}(\lambda) \leq t\} \quad \text{for} \quad t > 0.$$

Then $f^* : (0,\infty) \to (0,\infty)$ is decreasing, right-continuous and equimeasurable with $|f|$ (i.e., $d_{f^*} = d_{|f|}$). The function f^* is called the **decreasing rearrangement** of $|f|$ (see e.g., [**BS**]). In particular we recall that

$$\int_0^t f^*(s)ds = \sup\left\{ \int_A |f|dm : A \subset (0,\infty) \text{ measurable and } m(A) \leq t \right\}$$

(by [**BS**, Prop. 3.3., p. 53]).

Suppose that E is a linear subspace of $\mathcal{S}_0(0,\infty)$, which is a Banach space with respect to the norm $\|\cdot\|_E$. Then E will be called a **rearrangement invariant Banach function space** if

$$f \in E\,,\ g \in \mathcal{S}_0(0,\infty) \ \text{ and } \ g^* \leq f^* \ \text{ imply that } \ g \in E \ \text{ and } \ \|g\|_E \leq \|f\|_E$$

(see e.g., [**KPS**]). If E is such a rearrangement invariant space on $(0,\infty)$, we always have the continuous embeddings

$$L^1 \cap L^\infty(0,\infty) \subseteq E \subseteq (L^1 + L^\infty)(0,\infty).$$

Here the spaces $L^1 \cap L^\infty$ and $L^1 + L^\infty$ are equipped with the norms

$$\begin{aligned} \|f\|_{L^1\cap L^\infty} &= \max\{\|f\|_1\,,\ \|f\|_\infty\}, \\ \|f\|_{L^1+L^\infty} &= \inf\{\|g\|_1 + \|h\|_\infty : f = g + h, \\ &\qquad g \in L^1(0,\infty),\ h \in L^\infty(0,\infty)\}, \end{aligned}$$

respectively.

Given f, $g \in \mathcal{S}_0(0,\infty)$, we say that g is **submajorized** by f (in the sense of Hardy-Littlewood-Polya) if

$$\int_0^t g^*(s)ds \leq \int_0^t f^*(s)ds \quad \text{for all} \quad t > 0,$$

which is denoted by $g \prec\prec f$.

Using this submajorization relation the exact (L^1, L^∞)-interpolation spaces can be characterized. In fact, it is a result of A.P. Calderon (e.g., see [**BS**, Theorem 2.12]) that a Banach space $(E, \|\ \|_E)$, with $E \subseteq (L^1 + L^\infty)(0,\infty)$, is an exact (L^1, L^∞)-interpolation space if and only if,

$$f \in E,\ g \in \mathcal{S}_0(0,\infty) \ \text{ and } \ g \prec\prec f \ \text{ imply that } \ g \in E \ \text{ and } \ \|g\|_E \leq \|f\|_E\,.$$

In particular, such interpolation spaces are rearrangement invariant Banach function spaces. Although some of the results in this paper hold for more general rearrangement invariant spaces, we will assume that the spaces we consider are exact (L^1, L^∞)-interpolation spaces. This class includes many of the classical function spaces (e.g., L^p-spaces, Orlicz spaces, Lorenz spaces, Marcinkiewiecz spaces).

If E is a rearrangement invariant Banach function space on $(0,\infty)$ which is **monotone complete** (i.e., $0 \le f_n \in E$, $f_n \le f_{n+1}$ *a.e.*, $\sup_n \|f_n\|_E < \infty$ implies that there exists $0 \le f \in E$ such that $f_n \uparrow f$ *a.e.* and $\|f\|_E = \sup_n \|f_n\|_E$), then E is an exact (L^1, L^∞)-interpolation space (see e.g., [**BS**, Theorem 2.2, p. 106]).

Similarly, any rearrangement invariant Banach function space with order continuous norm is an exact (L^1, L^∞)-interpolation space.

Since every interpolation space can be renormed in such a way that it becomes an exact interpolation space (see [**BS**]), in the following we will assume that the interpolation space is exact, throughout the paper.

For $s > 0$ the dilation operator D_s, acting on measurable functions f on $(0,\infty)$, is defined by

$$D_s f(t) = f(t/s), \ t > 0.$$

Clearly, the operators D_s are bounded on any (L^1, L^∞)-interpolation space E and satisfy $\|D_s\|_E \le \max(1, s)$ for all $s > 0$. Note that $(D_s f)^* = D_s f^*$ for all $s > 0$ and all $f \in E$, so in particular $\|D_s f\|_E$ is an increasing function of s.

For such a space E the **upper and lower Boyd indices** are defined by

$$\overline{\alpha}_E = \lim_{s \to \infty} \frac{\log \|D_s\|}{\log s}, \ \underline{\alpha}_E = \lim_{s \downarrow 0} \frac{\log \|D_s\|}{\log s}$$

respectively, and satisfy $0 \le \underline{\alpha}_E \le \overline{\alpha}_E \le 1$ (see e.g., [**BS**], [**KPS**]). By way of example, if $E = L^p \cap L^q(0,\infty)$, $1 \le p \le q \le \infty$, (equipped with the norm $\|f\|_E = \max(\|f\|_p, \|f\|_q)$), then $\underline{\alpha}_E = 1/q$, $\overline{\alpha}_E = 1/p$.

In Section 4 we will use the following result.

Lemma 1.1. *Let E be an (L^1, L^∞)-interpolation space on $(0,\infty)$ and μ a (positive) Borel measure on $(0,\infty)$.*

Suppose that $f \in E$ satisfies $\int_0^\infty \|D_s f\|_E \, d\mu(s) < \infty$.

Then $\int_0^\infty D_s f(x) d\mu(s)$ is absolutely convergent for almost all $x > 0$,

$$\int_0^\infty D_s f(\cdot) d\mu(s) \in E \quad \text{and} \quad \left\| \int_0^\infty D_s f(\cdot) d\mu(s) \right\|_E \le \int_0^\infty \|D_s f\|_E \, d\mu(s).$$

In particular, if $\int_0^\infty \|D_s\|_E \, d\mu(s) < \infty$, then

$$T_\mu f(x) = \int_0^\infty D_s f(x) d\mu(s), \ \textbf{a.e.} \ \ x \in (0,\infty),$$

defines a bounded linear operator in E satisfying

$$\|T_\mu\|_E \le \int_0^\infty \|D_s\|_E \, d\mu(s).$$

Proof. The proof is divided in two parts.

1. Suppose that $f \in (L^1 + L^\infty)(0,\infty)$ is such that $\int_0^\infty D_s f^*(\cdot) d\mu(s) \in (L^1 + L^\infty)(0,\infty)$. We claim that $\int_0^\infty D_s f(x) d\mu(s)$ is absolutely convergent for a.e. $x \in (0,\infty)$, and that

$$\int_0^\infty D_s f(\cdot) d\mu(s) \prec\!\!\prec \int_0^\infty D_s f^*(\cdot) d\mu(s).$$

Indeed, for any measurable set $A \subseteq (0,\infty)$ with $m(A) < \infty$ we have

$$\int_A \left(\int_0^\infty |D_s f(x)| d\mu(s) \right) dx = \int_0^\infty \left(\int_A D_s f(x) dx \right) d\mu(s) \le$$

$$\int_0^\infty \left(\int_0^{m(A)} D_s f^*(x) dx \right) d\mu(s) = \int_0^{m(A)} \left(\int_0^\infty |D_s f^*(x)| d\mu(s) \right) dx < \infty.$$

This shows in particular that $\int_0^\infty |D_s f(x)| d\mu(s) < \infty$ for a.e. $x \in (0,\infty)$. Moreover,

$$\begin{aligned} &\int_0^t \left(\int_0^\infty |D_s f(\cdot)| d\mu(s) \right)^* (x) dx \\ &= \sup \left\{ \int_A \left(\int_0^\infty |D_s f(x)| d\mu(s) \right) dx \,:\, m(A) \le t \right\} \\ &\le \int_0^t \left(\int_0^\infty D_s f^*(x) d\mu(s) \right) dx \quad \text{for all} \quad t > 0, \end{aligned}$$

and since $|\int_0^\infty D_s f(x) d\mu(s)| \le \int_0^\infty |D_s f(x)| d\mu(s)$ the claim follows.

2. Now assume that $f \in E$ is such that $\int_0^\infty ||D_s f||_E d\mu(s) < \infty$. Since $||D_s f||_E = ||D_s f^*||_E$, it follows from [**KPS**, II.4.7] that $\int_0^\infty D_s f^*(\cdot) d\mu(s) \in E$ and

$$\left\| \int_0^\infty D_s f^*(\cdot) d\mu(s) \right\|_E \le \int_0^\infty \|D_s f\|_E \, d\mu(s).$$

From 1. above it follows that $\int_0^\infty D_s f(x) d\mu(s)$ is absolutely convergent for a.e. $x \in (0,\infty)$, and

$$\int_0^\infty D_s f(\cdot) d\mu(s) \prec\!\!\prec \int_0^\infty D_s f^*(\cdot) d\mu(s).$$

Since E is an exact (L^1, L^∞)-interpolation space, this implies that $\int_0^\infty D_s f(\cdot) d\mu(s) \in E$ and

$$\left\| \int_0^\infty D_s f^*(\cdot) d\mu(s) \right\|_E \leq \int_0^\infty \|D_s f\| D_s f \|_E \, d\mu(s).$$

Finally it should be observed that the function $\int_0^\infty D_s f(\cdot) d\mu(s)$ does not depend (modulo Lebesgue null sets) on the choice of the representative f. □

Next we recall some notions and results concerning resolvent positive operators which will be needed later. Let E be a Banach lattice. An operator A on E is called **resolvent positive** if there exists a number $\lambda_0 \in \mathbb{R}$ such that $(\lambda_0, \infty) \subset \varrho(A)$ and $R(\lambda, A) \geq 0$ for all $\lambda > \lambda_0$. Denote by

$$s(A) = \sup\{Re\, \lambda : \lambda \in \sigma(A)\}$$

the **spectral bound** of A. It is known that

$$s(A) = \inf\{\lambda \in \mathbb{R} \cap \varrho(A) : R(\lambda, A) \geq 0\}$$

and that, $s(A) \in \sigma(A)$ if $s(A) > -\infty$. Moreover, one hat

$$0 \leq R(\mu, A) \leq R(\lambda, A) \text{ if } \mu > \lambda > s(A) \tag{1.1}$$

and

$$|R(\lambda, A)x| \leq R(Re\, \lambda, A)|x| \tag{1.2}$$

for all $x \in E$, $Re\ \lambda > s(A)$. We say that A **generates an integrated semigroup**, if there exists a stronlgy continuous increasing function $S : [0, \infty) \to \mathcal{L}(E)$ satisfying $S(0) = 0$ such that

$$R(\lambda, A) = \int_0^\infty e^{-\lambda t}\ dS(t) \qquad (\lambda > \lambda_0) \tag{1.3}$$

(as an improper strongly defined Riemann-Stieltjes integral) for some $\lambda_0 \geq \varrho(A)$. In that case S is called **the integrated semigroup generated by** A, and it is known that (1.3) converges whenever $Re\, \lambda > s(A)$. Moreover,

$$R(\lambda, A) = \lambda \int_0^\infty e^{-\lambda t} S(t)\ dt \qquad (Re\, \lambda > \max\{s(A), 0\}). \tag{1.4}$$

We need the following lemma.

Lemma 1.2. *Assume that S is bounded. Then $s(A) < 0$.*

Proof. It follows from [**A2**] Proposition 6.1 that $s(A) \leq 0$. Now (1.4) implies that $\|R(\lambda, A)\| \leq M = \sup_{t \geq 0} \|S(t)\|$ for $\lambda > 0$. This implies that $0 \in \varrho(A)$ and $R(0, A) = \lim_{\lambda \downarrow 0} R(\lambda, A) \geq 0$. Then for small $\mu < 0$

$$R(\mu, A) = \sum_{n=0}^\infty (-\mu)^n R(0, A)^{n+1} \geq 0.$$

This implies that $s(A) < 0$. □

It is known that a resolvent positive operator generates a once integrated semigroup if $D(A)$ is dense or E has order continuous norm. We refer to [**A2**] for this and further information. Without any further assumption, it is known ([**A3**, Corollary 4.5]) that every resolvent positive operator A generates a twice integrated semigroup S_2; i.e., $S_2 : [0, \infty) \to \mathcal{L}(E)$ is strongly continuous increasing function such that

$$R(\lambda, A) = \int_0^\infty \lambda^2 e^{-\lambda t} S_2(t)\; dt \qquad (Re\,\lambda > \max\{s(A), 0\}).$$

Of course, if A generates a C_0-semigroup, then $S(t) = \int_0^t T(s)\; ds$ is the once-integrated semigroup and $S_2(t) = \int_0^t \int_0^s T(r)dr\; ds$ the twice integrated semigroup generated by A.

2. The Cesaro operator in spaces with order continuous norm.

In this section we will show that the theory of strongly continuous positive semigroups provides on efficient framework to compute the spectrum of the Cesaro operator in certain rearrangement invariant Banach function spaces. Let E be an $(L^1, L^\infty)-$interpolation space on $(0, \infty)$. For $t \in \mathbb{R}$ and $f \in E$ let $T(t)f(x) = f(e^{-t}x)$ for a.e. $x \in (0, \infty)$. This defines a bounded linear operator $T(t)$ on E satisfying $\|T(t)\|_E \leq \max(1, e^t)$, and $\mathcal{T}_E = \{T(t)\}_{t\in\mathbb{R}}$ is a group. The growth bounds of this group are

$$\omega_0^+(\mathcal{T}_E) \;:=\; \lim_{t\to\infty} \frac{\log \|T(t)\|_E}{t} \;=\; \lim_{s\to\infty} \frac{\log \|D_s\|}{\log s} = \overline{\alpha}_E,$$

$$\omega_0^-(\mathcal{T}_E) \;:=\; \lim_{t\to\infty} \frac{\log \|T(-t)\|_E}{t} \;=\; \lim_{s\downarrow 0} \frac{\log \|D_s\|}{\log s} = -\underline{\alpha}_E.$$

Now assume in addition that E has order continuous norm. Using that stepfunctions on bounded intervals are dense in E, it follows immediately that $\mathcal{T}_E$ is a strongly continous group. Let A_E be the generator of $\mathcal{T}_E$. The spectral bounds of A_E are defined by

$$s^+(A_E) = \sup\{Re\,\lambda \,:\, \lambda \in \sigma(A_E)\},$$

$$s^-(A_E) = \sup\{Re\,\lambda \,:\, \lambda \in \sigma(-A_E)\} \;=\; s^+(-\,A_E).$$

Then $-\,\omega_0^-(\mathcal{T}_E) \;\leq\; -s^-(A_E) \;\leq\; s^+(A_E) \;\leq\; \omega_0^+(A_E)$, and

$$\sigma(A_E) \subseteq \{\lambda \in \mathbb{C} : -s^-(A_E) \leq Re\,\lambda \leq s^+(A_E)\}.$$

Theorem 2.1. *Let E be a rearrangement invariant Banach function space on $(0, \infty)$ with order continuous norm. Then*

$$\sigma(A_E) \;=\; \{\lambda \in \mathbb{C} \,:\, \underline{\alpha}_E \;\leq\; Re\,\lambda \;\leq\; \overline{\alpha}_E\}.$$

Proof. The proof is divided in three steps:

1. First we show that $\sigma(A_E)$ is invariant under purely imaginary translations. To this end, for $\tau \in \mathbb{R}$ we define the isometry $M_\tau : E \to E$ by $M_\tau f(x) = x^{i\tau} f(x)$ for a.e. $x \in (0,\infty)$ and all $f \in E$. Then $M_\tau^{-1} T(t) M_\tau = e^{-it\tau} T(t)$ for all $t, \tau \in \mathbb{R}$, and so $M_\tau^{-1} A_E M_\tau = A_E - i\tau$ for all $\tau \in \mathbb{R}$.
Hence $\sigma(A_E) = \sigma(M_\tau^{-1} A_E M_\tau) = \sigma(A_E) - i\tau$ for all $\tau \in \mathbb{R}$.

2. Next we will show that $\sigma(A_E) \cap \mathbb{R} = [-s^-(A_E), s^+(A_E)]$. It is clear that $\sigma(A_E) \cap \mathbb{R} \subseteq [-s^-(A_E), s^+(A_E)]$. Moreover, since $\mathcal{T}_E$ consists of positive operators, $s^+(A_E), -s^-(A_E) \in \sigma(A_E)$ (see e.g., [**N**], C - III, Theorem 1.1). Take $\mu \in \rho(A_E) \cap \mathbb{R}$. We claim that either $\mu > s^+(A_E)$ or $\mu < -s^-(A_E)$. Indeed, defining

$$\begin{aligned} I_\mu &= \{f \in E : R(\mu, A_E)|f| \geq 0\} \quad \text{and} \\ J_\mu &= \{f \in E : R(\mu, A_E)|f| \leq 0\}, \end{aligned}$$

it follows from the Theorem on p. 43 in [**G**] that I_μ and J_μ are $\mathcal{T}_E$−invariant bands satisfying $E = I_\mu \oplus J_\mu$. Since any band in E is of the form $\{f \in E : f = 0 \text{ a.e. on } B\}$ for some measurable subset $B \subseteq (0,\infty)$, it is easy to see that the only $\mathcal{T}_E$−invariant bands are E and $\{0\}$. Hence $I_\mu = E$ or $J_\mu = E$. Suppose that $I_\mu = E$. From the definition of I_μ it then follows that $R(\mu, A_E) \geq 0$, which implies that $\mu > s^+(A_E)$ (see [**N**, C - III, Theorem 1.1.]). If $J_\mu = E$, a similar argument shows that $\mu < -s^-(A_E)$, by which the claim is proved.

3. Finally we show that $s^+(A_E) = \overline{\alpha}_E$ and $s^-(A_E) = -\underline{\alpha}_E$. Take $\lambda > s^+(A_E)$. Then (see e.g., [**N**, C - III, Theorem 1.2.])

$$R(\lambda, A_E) f = \int_0^\infty e^{-\lambda t} T(t) f \, dt \quad \text{for all} \quad f \in E.$$

Fix $f \in E$. Observe that $(T(t)f)^* = T(t)f^*$ and that the function $t \mapsto T(t)f^*$ is increasing for $t \geq 0$. Hence

$$R(\lambda, A_E) f^* = \int_0^\infty e^{-\lambda s} T(s) f^* \, ds \geq \int_t^\infty e^{-\lambda s} T(s) f^* \, ds \geq \frac{e^{-\lambda t}}{\lambda} T(t) f^*$$

for all $t \geq 0$ (note that $\lambda > s^+(A_E) \geq -\omega_0^-(\mathcal{T}_E) = \underline{\alpha}_E \geq 0$). This implies that

$$\|T(t)f\|_E \leq \lambda e^{\lambda t} \|R(\lambda, A_E) f^*\|_E \leq \lambda e^{\lambda t} \|R(\lambda, A_E)\|_E \, ||f||_E$$

for all $t \geq 0$. This shows that $\omega_0^+(\mathcal{T}_E) \leq \lambda$, and consequently $\omega_0^+(\mathcal{T}_E) \leq s^+(A_E)$. Hence $\omega_0^+(\mathcal{T}_E) = s^+(A_E)$, i.e., $\overline{\alpha}_E = s^+(A_E)$. Via a similar argument it follows that $s^-(A_E) = -\underline{\alpha}_E$. Combining the results of (1), (2) and (3) we see that

$$\sigma(A_E) = \{\lambda \in \mathbb{C} : \underline{\alpha}_E \leq Re\, \lambda \leq \overline{\alpha}_E\}.$$

□

Recall that the **Cesaro operator** C on $(0,\infty)$ is given by

$$Cf(x) = \frac{1}{x}\int_0^x f(u)du\,, \qquad x > 0,$$

defined for functions f on $(0,\infty)$ which are integrable on $(0,x)$ for all $x > 0$. If E is a rearrangement invariant Banach function space on $(0,\infty)$ such that $Cf \in E$ for all $f \in E$, we denote the induced operator in E by C_E. Then C_E is a positive, and so a bounded operator on E.

Corollary 2.2. *Let E be a rearrangement invariant Banach function space on $(0,\infty)$ with order continuous norm. Then the Cesaro operator is bounded on E if and only if $\overline{\alpha}_E < 1$. In that case the spectrum $\sigma(C_E)$ of C_E is given by*

$$\sigma(C_E) = \left\{\lambda \in \mathbb{C} : 1 - \overline{\alpha}_E \le Re\left(\frac{1}{\lambda}\right) \le 1 - \underline{\alpha}_E\right\} \cup \{0\}.$$

Proof. Assume that $\overline{\alpha}_E < 1$. Then $s(A_E) < 1$ by Theorem 2.1. Moreover, we have

$$\begin{aligned}(R(1,A_E)f)(x) &= \left(\int_0^\infty e^{-t}T(t)f\,dt\right)(x) \\ &= \int_0^\infty e^{-t}f(e^{-t}x)\,dt = \frac{1}{x}\int_0^x f(u)du\end{aligned}$$

for almost all $x \in E$ and all $f \in E$.

Conversely, assume that the Cesaro operator is bounded on E. Consider the operators $S(t) = \int_0^t e^{-s}T(s)\,ds$. Then

$$(S(t)f)(x) = \frac{1}{x}\int_{e^{-t}x}^x f(u)du \le (C_E f)(x) \ a.e.$$

Hence $\|S(t)\| \le \|C_E\| \qquad (t \ge 0)$. This implies that $s(A_E) < 1$ by Lemma 1.2.

Now assume that $\overline{\alpha}_E < 1$. Then $C_E = R(1,A_E)$. From the spectral mapping theorem for resolvents and Theorem 2.1. it follows that

$$\begin{aligned}\sigma(C_E) &= \left\{\frac{1}{1-z} : z \in \sigma(A_E)\right\} \cup \{0\} \\ &= \left\{\lambda \in \mathbb{C} : 1 - \overline{\alpha}_E \le Re\left(\frac{1}{\lambda}\right) \le 1 - \underline{\alpha}_E\right\} \cup \{0\}.\end{aligned}$$

□

Remark 2.3. 1. For $E = L^p(0,\infty)$, $1 < p < \infty$, the result of the above corollary was obtained by D.W. Boyd in [**Bo**]. In the same paper the result of the above corollary is announced but, as fas as we know, a proof was never published.

2. For a large class of rearrangement invariant Banach function spaces E on $(0,\infty)$ it is well-known that boundedness of C_E is equivalent with $\overline{\alpha}_E < 1$. For spaces E with the Fatou property, this result is originally due to D.W. Boyd [**Bo**]. Proofs can also be found in e.g., [**BS**]. In a later section of the present paper we will discuss this equivalence for general (L^1, L^∞)-interpolation spaces.

3. In [**A4**] the semigroup T_E has been used on $E = L^p(0,\infty)$ to produce an example of p-dependent spectrum. It is remarkable that on $L^p \cap L^q(1,\infty)$, $p \neq q$, the type of the semigroup is strictly larger than the spectral bound [**A5**].

4. In the proof of Theorem 2.1 it was not necessary to compute the explicit form of the generator A_E of $\mathcal{T}_E$. However, it is not difficult to show that this generator is given by $A_E f(x) = -xf'(x)$, a.e. $x \in (0,\infty)$, with domain

$$D(A_E) = \{f \in E : f \in AC_{\text{loc}}(0,\infty) \quad \text{and} \quad xf'(x) \in E\}.$$

We leave the details to the reader.

Crucial in the above approach is the strong continuity of the group $\mathcal{T}_E$. As we have seen, if E has order continuous norm, then $\mathcal{T}_E$ is strongly continuous. We will show next that strong continuity of $\mathcal{T}_E$ implies that E has order continuous norm. In the theorem which follows we need not assume that E is an (L^1, L^∞)-interpolation space. In fact, if E is any rearrangement invariant Banach function space on $(0,\infty)$, then $\mathcal{T}_E$ is a group of bounded linear operators in E with $\|T(t)\| \leq \max(1, e^t)$ for all $t \in \mathbb{R}$ (this follows from [**KPS**, Section II, 4.3]).

Theorem 2.4. *Let E be a rearrangement invariant Banach function space on $(0,\infty)$. The group $\mathcal{T}_E$ is strongly continuous if and only if E has order continuous norm.*

In the proof of this theorem we will use a criterion for order continuity of the norm which is implicit in [**KPS**, (Section II, 4.5)]. For the sake of convenience we will state this criterion in the next lemma and provide the proof.

Lemma 2.5. *Let E be a rearrangement invariant Banach function space on $(0,\infty)$. Then E has order continuous norm if and only if*

(i) $\left\|f^*\chi_{(0,\frac{1}{n})}\right\|_E \to 0 \ (n \to \infty)$ *for all* $f \in E$;

(ii) $\left\|f^*\chi_{(n,\infty)}\right\|_E \to 0 \ (n \to \infty)$ *for all* $f \in E$.

Proof. It is clear that order continuity of the norm implies (i) and (ii). Now assume that E satisfies (i) and (ii). First observe that (ii) implies that $f^*(t) \to 0$ as $t \to \infty$ for all $f \in E$, i.e., that $m\{x \in (0,\infty) : |f(x)| > \lambda\} <$

∞ for all $\lambda > 0$ and all $f \in E$. Now suppose that $f_n \in E \quad (n = 1, 2, \dots)$ such that $f_n \downarrow 0$ a.e.. Let $\varepsilon > 0$ be given. By (i), (ii) there exists $N \in \mathbb{N}$ such that $\left\| f_1^* \chi_{(0,1/N)} \right\|_E < \varepsilon$ and $\left\| f_1^* \chi_{(N,\infty)} \right\|_E < \varepsilon$. From the above observation it follows that $f_n^*(1/N) \downarrow 0$ as $n \to \infty$. Hence there exists $n_0 \in \mathbb{N}$ such that $f_n^*(1/N) < \varepsilon$ for all $n \geq n_0$. For $n \geq n_0$ we have

$$\begin{aligned} \|f_n\|_E = \|f_n^*\|_E &\leq \left\| f_n^* \chi_{(0,\frac{1}{N})} \right\|_E + \left\| f_n^* \chi_{[\frac{1}{N},N]} \right\|_E + \left\| f_n^* \chi_{(N,\infty)} \right\|_E \\ &\leq \left\| f_1^* \chi_{(0,1/N)} \right\|_E + \varepsilon \left\| \chi_{[1/N,N]} \right\|_E + \left\| f_1^* \chi_{(N,\infty)} \right\|_E \\ &\leq 2\varepsilon + \varepsilon C \left\| \chi_{[1/N,N]} \right\|_{L^1+L^\infty} \leq (2 + C)\varepsilon, \end{aligned}$$

where $C > 0$ is the embedding constant of E into $(L^1 + L^\infty)(0, \infty)$. This shows that $\|f_n\|_E \downarrow 0$ $(n \to \infty)$, and we may conclude that E has order continuous norm. □

Proof of Theorem 2.4. As observed already above, if E has order continuous norm, then $\mathcal{T}_E$ is strongly continuous. Now assume that $\mathcal{T}_E$ is strongly continuous. Fix $f \in E$ and define $g(s) = f^*(s-1)$ for $s > 1$ and $g(s) = 0$ for $0 < s \leq 1$. Then $g^* = f^*$, so $g \in E$. Since $T(t)g(s) = 0$ for $0 < s \leq e^t$, it follows that $|T(t)g - g| \geq g\chi_{(1,e^t]}$ for all $t > 0$, and so $\left\| g\chi_{(1,e^t]} \right\|_E \leq \|T(t)g - g\|_E$ for all $t > 0$. Hence $\left\| g\chi_{(1,e^t]} \right\|_E \to 0$ as $t \downarrow 0$. Now $(g\chi_{(1,e^t]})^* = f^*\chi_{(0,e^t-1]}$ implies that $\left\| f^*\chi_{(0,e^t-1]} \right\|_E \to 0$ as $t \downarrow 0$, which shows that $\left\| f^*\chi_{(0,1/n)} \right\|_E \to 0 \quad (n \to \infty)$. It remains to show that $\left\| f^*\chi_{(n,\infty)} \right\|_E \to 0$ $(n \to \infty)$. Define $n_0 = 0$ and $n_k = 3(n_{k-1} + 1)$ for $k = 1, 2, \dots,$ and let

$$h(s) = \begin{cases} f(s + k - n_k - 1) & \text{if } n_k < s \leq n_k + 1\,,\ k = 1, 2 \dots \\ 0 & \text{otherwise.} \end{cases}$$

Then $h^* = f^*$, so $f \in E$. Now let $\varepsilon > 0$ be given. By the strong continuity of $\mathcal{T}_E$, there exists $0 < t_0 \leq 1$ such that $\|T(-t_0)h - h\|_E < \varepsilon$. Take k_0 such that $e^{-t_0} < n_k\ (n_k + 1)^{-1}$ for all $k \geq k_0$. Now observe that h is supported on the set $\bigcup_{k=1}^{\infty} (n_k\,,\ n_k + 1]$ and $T(-t_0)h$ is supported on $\bigcup_{k=1}^{\infty} (e^{-t_0}n_k\,,\ e^{-t_0}(n_k + 1)]$. Since, by the definition of the n_k's and by the choice of k_0 , $n_{k-1} + 1 < e^{-t_0}n_k < e^{-t_0}(n_k + 1) < n_k$ for all $k \geq k_0$, it follows that $h\chi_{(n_{k_0},\infty)}$ and $T(-t_0)\chi_{(n_{k_0},\infty)}$ are disjointly supported. Hence,

$$|T(-t_0)h - h| \geq |T(-t_0)h - h|\chi_{(n_{k_0},\infty)} \geq h\chi_{(n_{k_0},\infty)}$$

which implies that $\left\|h\chi_{(n_{k_0},\infty)}\right\|_E < \varepsilon$. Now

$$(h\chi_{(n_{k_0},\infty)})^* = f^*\chi_{(k_0-1,\infty)}\,, \text{ so } \left\|f^*\chi_{(k_0-1,\infty)}\right\|_E < \varepsilon.$$

This shows that $\left\|f^*\chi_{(n,\infty)}\right\|_E \to 0$ $(n \to \infty)$. Via Lemma 2.5 it now follows that E has order continuous norm. □

The above theorem shows in particular that it is not possible to compute the spectrum of the Cesaro operator using the theory of strongly continuous (semi)groups, as in the proof of Theorem 2.1, if the space E does not have order continuous norm. This is one of the motivations for the investigations in the next section. In particular we will need an appropriate substitute for the spectral decomposition theorem for generators of strongly continuous groups of G. Greiner [**G**].

3. Spectral decomposition.

Througout this section we assume that A is an operator on a complex Banach lattice E such that $\pm A$ is resolvent positive (we say that A is **resolvent bipositive**). Then we know from the proof of [**N**, C-III Corollary 1.6] that $\sigma(A) \neq \emptyset$. Denote by

$$s(A) = \sup\{Re\ \lambda : \lambda \in \sigma(A)\}$$

the **spectral bound** of A. Then we know that

(3.1) $\sigma(A) \subset \{\lambda \in \mathbb{C} : -s(-A) \leq Re\ \lambda \leq s(A)\}$;

(3.2) $s(A), -s(-A) \in \sigma(A)$;

(3.3) $R(\lambda, A) \geq 0$ if $\lambda > s(A)$;

(3.4) $R(\lambda, A) \leq 0$ if $\lambda < -s(-A)$.

Definition 3.1. Let $\mu \in (-s(-A),\ s(A))$. We say that A **allows a spectral decomposition with respect to μ** if there exists a band decomposition $E = E_1 \oplus E_2$ such that $R(\lambda, A)E_i = E_i \quad (i = 1, 2)$ for all $\lambda \in \varrho(A)$ and such that the part A_i of A in E_i satisfies

$$\sigma(A_1) = \{\lambda \in \sigma(A) : Re\,\lambda < \mu\},$$
$$\sigma(A_2) = \{\lambda \in \sigma(A) : Re\,\lambda > \mu\}.$$

In particular, in that case $\pm A_i$ is resolvent positive and $s(A_1) < \mu$ and $-s(-A_2) > \mu$.

The main result of this section is the following:

Theorem 3.2. *Let $\mu \in (-s(-A),\ s(A)) \cap \varrho(A)$. Then A allows a spectral decomposition with respect to μ if one of the following two conditions is satisfied.*

(a) *The operator* A *satisfies*

(K_μ) $x \in D(A)$ *implies* $|x| \in D(A)$ *and* $|(\mu - A)|x|\,| \leq |(\mu - A)x|$;

(b) *The domain* $D(A)$ *is dense.*

The condition (K_μ) is a weak form of Kato's equality which we will discuss later.

For the proof we can assume that $\mu = 0$ which we will do in the following. It is known that for $\lambda \in \varrho(A) \cap \mathbb{R}$ one has $R(\lambda, A) \geq 0$ if and only if $\lambda > s(A)$ (see [**N**, C-III Theorem 1.1]). In view of this, following Greiner's idea [**N**, C-III Theorem 4.8], we set

$$E_1 = \{x \in E : R(0, A)|x| \geq 0\};$$
$$E_2 = \{x \in E : R(0, A)|x| \leq 0\}.$$

Lemma 3.3. a) *E_1 and E_2 are closed ideals in E.*

b) *The operator A allows a spectral decomposition with respect to* 0 *whenever $E_1 + E_2 = E$.*

Remark 3.4. Lemma 3.3 a) is true without additional hypotheses. Conditions (a), (b) of Theorem 3.2 are used to show that $E = E_1 + E_2$. Our point is to replace in Greiner's argument the semigroup (which does not need to exist here, see Example 3.13) by the twice integrated semigroup. Moreover, we simplify the argument using the following description of the abscissa of the Laplace transform (see [**ANS**, Proposition 1.1] or [**HP**, Sec. 6.2] for a proof).

Lemma 3.5. *Let X be a Banach space and $f : [0, \infty) \to X$ be continuous. Then abs $(f) \leq 0$ (i.e., $\hat{f}(\lambda) := \lim_{t\to\infty} \int_0^t e^{-\lambda s} f(s)\, ds$ exists whenever $Re\, \lambda > 0$) if and only if*

$$\sup_{t \geq 0} e^{-wt} \left\| \int_0^t f(s)\ ds \right\| < \infty, \quad \textit{for all} \ \ w > 0.$$

Proof of Lemma 3.3. Let S be the twice integrated semigroup generated by A; that is, $S : [0, \infty) \to \mathcal{L}(E)$ is strongly continuous and there exists $\omega \geq 0$ such that $(\omega, \infty) \subset \varrho(A)$, $\sup_{t \geq 0} \|e^{-\omega t} S(t)\| < \infty$ and,

$$R(\lambda, A)x = \lambda^2 \int_0^\infty e^{-\lambda t} S(t)x\ dt \quad (Re\, \lambda > \omega,\ x \in E).$$

Then $S(t) \geq 0$ and $S(t)R(\lambda, A) = R(\lambda, A)S(t)$ for all $\lambda \in \varrho(A)$, $t \geq 0$ and for all $x \in E$,

$$\text{(3.5)} \qquad \int_0^t S(s)x\ ds \in D(A) \text{ and } A \int_0^t S(s)x\ ds = S(t)x - \frac{t^2}{2}x.$$

See Section 1 and [**A3**] for these notions and results. We claim that for $x \in E$,

(3.6) $$x \in E_1 \text{ if and only if } \; abs\ (S(\cdot)|x|) \leq 0.$$

In fact, if $x \in E_1$, then by (3.5) (since $0 \in \varrho(A)$),

$$\begin{aligned} \int_0^t S(s)|x|\ ds &= \frac{t^2}{2}R(0,A)|x| - S(t)R(0,A)|x| \\ &\leq \frac{t^2}{2}R(0,A)|x| \qquad (t \geq 0). \end{aligned}$$

Hence $abs\ (S(\cdot)|x|) \leq 0$ by Lemma 3.5. Conversely, assume $abs\ (S(\cdot)|x|) \leq 0$. Then $r(\lambda) = \lambda^2 \int_0^\infty e^{-\lambda t} S(t)|x|\ dt \quad (Re\ \lambda > 0)$ is holomorphic and for $\lambda > \omega$ one has $r(\lambda) = R(\lambda, A)|x|$, or equivalently, $\lambda A^{-1} r(\lambda) - r(\lambda) = A^{-1}|x|$. This remains true for $Re\ \lambda > 0$ by the uniqueness of holomorphic extensions. Hence $R(\lambda, A)|x| = r(\lambda) \geq 0$ for $\lambda \in (0, \varepsilon)$ where $\varepsilon > 0$ such that $(0, \varepsilon) \subset \varrho(A)$. This implies $R(0, A)|x| \geq 0$.

a) It follows from (3.6) and Lemma 3.5 that E_1 is an ideal. Closedness follows from the definition. Replacing A by $-A$ we see that also $E_2 = \{x : R(0, -A)|x| \geq 0\}$ is a closed ideal.

b) It is clear that $E_1 \cap E_2 = \{0\}$. Now assume that $E_1 + E_2 = E$. Then E_1 and E_2 are projection bands.

Let $\lambda_0 > s(A)$. Since $R(\lambda_0, A) \geq 0$ and $R(\lambda_0, A)S(t) = S(t)R(\lambda_0, A)$ it follows from (3.6) and Lemma 3.5 that $R(\lambda_0, A)E_i \subset E_i \quad (i = 1, 2)$. Hence $R(\lambda_0, A)P_1 = P_1 R(\lambda_0, A)$ where P_1 denotes the band projection onto E_1. It follows easily that $x \in D(A)$ implies $P_1 x \in D(A)$ and $AP_1 x = P_1 Ax$; and this in turn implies $R(\lambda, A)P_1 = P_1 R(\lambda, A)$ for all $\lambda \in \varrho(A)$. Thus $R(\lambda, A)E_i \subset E_i \quad (i = 1, 2)$ for all $\lambda \in \varrho(A)$. Hence $\varrho(A) = \varrho(A_1) \cap \varrho(A_2)$. Finally, by the first part of the proof, $Q(\lambda)x = \lambda^2 \int_0^\infty e^{-\lambda t} S(t)x\ dt$ exists for all $x \in E_1$ and $Re\,\lambda > 0$. Thus $Q(\lambda) \in \mathcal{L}(E_1)$, $A^{-1}Q(\lambda) = Q(\lambda)A^{-1}$ and $\lambda A^{-1}Q(\lambda)x - Q(\lambda)x = A^{-1}x$ if $Re\,\lambda > \omega$ and so for $Re\,\lambda > 0$ by holomorphy. This implies that $\lambda \in \varrho(A_1)$ and $Q(\lambda) = (\lambda - A_1)^{-1}$ if $Re\ \lambda > 0$. Similarly, $\{\lambda : Re\,\lambda < 0\} \subset \varrho(A_2)$. □

Lemma 3.6. *If* (K_0) *holds, then* $E = E_1 + E_2$.

Proof. Let $0 \leq x \in E$ and $y = R(0, A)x$. Then $|y| \in D(A)$ and $|A|y|\ | \leq |Ay| = x$. Thus $x_1 := \frac{1}{2}(x - A|y|) \geq 0$ and $x_2 := \frac{1}{2}(x + A|y|) \geq 0$. Moreover, $R(0, A)x_1 = \frac{1}{2}(R(0, A)x + |y|) = \frac{1}{2}(y + |y|) = y^+ \geq 0$. Thus $y_1 \in E_1$. Similarly, $R(0, A)x_2 = -y^- \leq 0$ so that $x_2 \in E_2$. Clearly, $x = x_1 + x_2$. We have shown that $E_+ \subset E_{1+} + E_{2+}$. This implies the claim. □

Now we prove Theorem 3.2. Under the hypothesis (a), the proof is complete. Case (b) follows from the following lemma.

Lemma 3.7. *Assume that $D(A)$ is dense. Then $E = E_1 + E_2$.*

Proof. We can assume that E is a real Banach lattice.

a) Let $x \in D(A^3)$, $\sigma = \text{sign}\, q(x) \in \mathcal{L}(E'')$ where $q : E \to E''$ is the canonical embedding. We show that

$$\langle \sigma Ax,\ \varphi\rangle = \langle |x|,\ A'\varphi\rangle \qquad (\varphi \in D(A'^3)). \tag{3.7}$$

In fact, it follows from the resolvent equation that $R(\lambda, A)$ is decreasing on $(s(A), \infty)$. Consequently, $\lambda R(\lambda, A)y = R(\lambda, A)Ay + y$ is bounded on $[s(A)+1, \infty)$ and so $R(\lambda, A)y \to 0 \quad (\lambda \to \infty)$ for $y \in D(A)$. This implies that $\lambda R(\lambda, A)y \to y \quad (\lambda \to \infty)$ if $y \in D(A^2)$. Finally,

$$\lambda^2 R(\lambda, A)y - \lambda y = \lambda R(\lambda, A)Ay \to Ay \qquad (\lambda \to \infty)$$

if $y \in D(A^3)$. For the same reason,

$$\lambda^2 R(\lambda, A)'\varphi - \lambda\varphi \to A'\varphi \qquad (\lambda \to \infty)$$

if $\varphi \in D(A'^3)$. Consequently, if $0 \le \varphi \in D(A'^3)$, then

$$\begin{aligned}
\langle \sigma Ax, \varphi\rangle &= \lim_{\lambda\to\infty} \langle \sigma(\lambda^2 R(\lambda, A)x - \lambda x)\ ,\ \varphi\rangle \\
&= \lim_{\lambda\to\infty} \langle \sigma(\lambda^2 R(\lambda, A)x) - \lambda |x|,\ \varphi\rangle \\
&\le \limsup_{\lambda\to\infty} \langle \lambda^2 R(\lambda, A)|x| - \lambda|x|,\ \varphi\rangle \\
&= \limsup_{\lambda\to\infty} \langle |x|,\ \lambda^2 R(\lambda, A)'\varphi - \lambda\varphi\rangle \\
&= \langle |x|,\ A'\varphi\rangle.
\end{aligned}$$

Replacing A by $-A$ gives (3.7) for $0 \le \varphi \in D(A'^3)$. Let $\mu_0 > s(A)$. Since $D(A'^3) = R(\mu_0, A)'^3 E' = D(A'^3) \cap E'_+ - D(A'^3) \cap E'_+$ we obtain (3.7) for all $\varphi \in D(A'^3)$.

b) Next we assume that $\mu = 0 \in \varrho(A)$ as before. Given $y \in D(A^2)$, we show that there exists $z'' \in E''$ such that $|z''| \le |y|$ and

$$|R(0, A)y| = R(0, A)''z''. \tag{3.8}$$

In fact, let $x = R(0, A)y$, $\sigma = \text{sign}\, q(x)$, $z'' = \sigma y$. Let $\psi \in D(A'^2)$, $\varphi = R(0, A)'\psi$. Then by (3.7),

$$\begin{aligned}
\langle R(0, A)''z'',\ \psi\rangle &= \langle z'',\ R(0, A)'\psi\rangle \\
&= \langle z'', \varphi\rangle = -\langle \sigma Ax,\ \varphi\rangle = -\langle |x|,\ A'\varphi\rangle \\
&= \langle |x|,\ \psi\rangle.
\end{aligned}$$

Since $D(A'^2) = (R(0, A)')^2 E'$ separates points, (3.8) follows.

c) Let $y \in D(A^2)_+$. Then $(R(0,A)y)^+ \in E_1$. In fact,

$$\begin{aligned}(R(0,A)y)^+ &= 1/2(|R(0,A)y| + R(0,A)y)\\ &= 1/2(R(0,A)''z'' + R(0,A)y)\\ &= R(0,A)''y_1''\end{aligned}$$

where $y_1'' = 1/2(y + z'') \geq 0$. It follows from (3.5) that

$$\left(\int_0^t S(s)\ ds\right)'' = t^2/2\ R(0,A)'' - S(t)''R(0,A)''.$$

Hence

$$\begin{aligned}\left(\int_0^t S(s)\ ds\right)'' y_1'' &= t^2/2\ R(0,A)''y_1'' - S(t)''R(0,A)''y_1''\\ &= t^2/2\ (R(0,A)y)^+ - S(t)(R(0,A)y)^+\\ &\leq t^2/2\ (R(0,A)y)^+.\end{aligned}$$

Hence

$$\begin{aligned}\left\|\int_0^t S(s)(R(0,A)y)^+\ ds\right\| &= \left\|\int_0^t S(s)R(0,A)''\ y_1''\ \ ds\right\|\\ &\leq t^2/2\ \|R(0,A)\|\ \|(R(0,A)y)^+\|.\end{aligned}$$

Thus abs $(S(\cdot)(R(0,A)y)^+) \leq 0$ by Lemma 3.5.

It follows from (3.6) that $(R(0,A)y)^+ \in E_1$.

d) Let $y \in D(A^2)_+$. Then, applying c) to $(-A)$ we have

$$(R(0,A)y)^- = (R(0,-A)y)^+ \in E_2.$$

Thus $R(0,A)y = (R(0,A)y)^+ - (R(0,A)y)^- \in E_1 + E_2$.

Since for $\mu > s(A)$, $D(A^2) = R(\mu,A)^2E = R(\mu,A)^2E_+ - R(\mu,A)^2E_+$ one has $D(A^2) = D(A^2)_+ - D(A^2)_+$. Thus $D(A^3) = R(0,A)D(A^2) \subset E_1 + E_2$. Consequently, $E = \overline{D(A^3)} \subset \overline{E_1 + E_2} = E_1 + E_2$, the sum of closed ideals being closed [**S**, III.1.2]. □

Let $T \in \mathcal{L}(E)$. A band B of E is called **reducing for** T if $TB \subset B$ and $TB^d \subset B^d$ (equivalently, T commutes with the band projection onto B).

Corollary 3.8. *Let A be an operator on E such that*

(a) *$\pm A$ is resolvent positive;*
(b) *$R(\lambda, A)$ has no nontrivial reducing band for some (equivalently all) $\lambda \in \varrho(A)$.*
(c) *$D(A)$ is dense or A satisfies (K_μ) for all $\mu \in \mathbb{R}$.*

Then $\sigma(A) \cap \mathbb{R} = [-s(-A),\ s(A)]$.

Next we give several comments concerning the inequality (K_μ). Greiner [**G**] (see also [**N**, C-III Section 4]) uses Kato's equality

$$A|x| = Re\,((\text{sign } \overline{x})Ax) \tag{K}$$

in his proof of the decomposition theorem. It holds for all $x \in D(A)$ if A is the generator of a positive C_0-group on a σ-order complete Banach lattice. In particular, $D(A)$ is a sublattice of E. Here, for $x \in E$, $\overline{x} = Re\ x - iIm\,x$ denotes the complex conjugate of x. Moreover, for $x \in E$, the operator sign $\overline{x} \in \mathcal{L}(E)$ is uniquely determined by the properties

$$\begin{aligned} (\text{sign } \overline{x})x &= |x| \\ |(\text{sign } \overline{x})y| &\leq |y| \qquad (y \in E) \\ (\text{sign } \overline{x})y &= 0 \quad \text{if } \ y \perp x. \end{aligned}$$

It is clear that $A - \mu$ satisfies (K) for all $\mu \in \mathbb{R}$ if A satisfies (K). Thus condition (K) implies condition (K_μ) for all $\mu \in \mathbb{R}$.

However, the converse is not true. In fact, in the following proposition we show that the adjoint A' of the generator A of a positive C_0-group always satisfies (K_μ) for all $\mu \in \mathbb{R}$. However, we show by an example that (K) may be violated.

Proposition 3.9. *Let B be the generator of a positive C_0-group T on a Banach lattice F and let $A = B'$ on $E = F'$. Then A satisfies (K_μ) for all $\mu \in \mathbb{R}$.*

Proof. We can assume $\mu = 0$. Recall that $D(B') = \text{Fav}\ (B') = \{\varphi \in F' : \limsup_{t\downarrow 0} 1/t\|T(t)'\varphi - \varphi\| < \infty\}$, see [**EN**, Chapter II.5.19] or [**CH**]. Let $\varphi \in D(B')$. Let $0 \leq x \in E$, $1 \geq t \geq 0$. Then

$$\begin{aligned} \langle |T(t)'\varphi - \varphi|,\ x\rangle &= \sup_{|y|\leq x} |\langle T(t)'\varphi - \varphi,\ y\rangle| \\ &= \sup_{|y|\leq x} \left| \int_0^t \langle B'\varphi, T(s)y\rangle\ ds \right| \\ &\leq \int_0^t \langle |B'\varphi|, T(s)x\rangle\ ds. \end{aligned}$$

It follows that

$$\begin{aligned} \left\langle \frac{1}{t}(T(t)'|\varphi| - |\varphi|),\ x \right\rangle &= \left\langle \frac{1}{t}(|T(t)'\varphi| - |\varphi|),\ x \right\rangle \\ &\leq \left\langle \frac{1}{t}|T(t)'\varphi - \varphi|,\ x \right\rangle \\ &\leq \frac{1}{t}\int_0^t \langle |B'\varphi|, T(s)x\rangle\ ds \\ &\leq \|B'\varphi\| \sup_{0<t\leq 1} \|T(s)x\|. \end{aligned}$$

Thus, $|\varphi| \in \text{Fav}\ (B') = D(B')$. Moreover,

$$\begin{aligned} \langle B'|\varphi|\ ,\ x\rangle &= \lim_{t\downarrow 0} \frac{1}{t}\langle T(t)'|\varphi| - |\varphi|\ ,\ x\rangle \\ &\leq \overline{\lim_{t\downarrow 0}}\ \frac{1}{t}\int_0^t \langle |B'\varphi|\ ,\ T(s)x\rangle\ ds \\ &\leq \langle |B'\varphi|,\ x\rangle. \end{aligned}$$

Hence $B'|\varphi| \leq |B'\varphi|$. □

Remark 3.10. It follows from Proposition 3.9 that Theorem 3.2 also holds if A is the adjoint of a generator B of a positive C_0-group. But of course, this can be directly seen by applying Theorem 3.2 to B.

Next we show that in the situation of Proposition 3.7 it can happen that $|B'|\varphi|\ | \neq |B'\varphi|$ for some $\varphi \in D(B')$; in particular, B' does not satisfy (K) in general.

Example 3.11. Consider in the space $E = C_0(\mathbb{R})$, equipped with the sup-norm, the C_0-group $(T(t))_{t\in\mathbb{R}}$ given by $T(t)f(x) = f(x+t)$ for all $x, t \in \mathbb{R}$. The generator B of this group is given by $Bf = f'$ with $D(B) = \{f \in C_0^1(\mathbb{R}) : f' \in C_0(\mathbb{R})\}$. Identifying the dual space $C_0(\mathbb{R})'$ with the space $M_b(\mathbb{R})$ of all bounded Borel measures on $\mathbb{R}$, it is easy to see that $D(B') = \{\mu \in M_b(\mathbb{R}) : D\mu \in M_b(\mathbb{R})\}$ and $B'\mu = -D\mu$ for all $\mu \in D(B')$, where $D\mu$ denotes the distributional derivate of the measure μ. As is well-known, every $\mu \in D(B')$ is absolutely continuous with respect to Lebesgue measure and is of the form $\mu = f dx$ with $f \in L^1(\mathbb{R}) \cap BV(\mathbb{R})$. Moreover, for such measures μ we have $D\mu = df$ (where df denotes the Borel measure induced by f). Now take $f = -1_{(-1,0]} + 1_{(0,1]}$ and $\mu = f dx$. Then $\mu \in D(B')$ and $B'\mu = -D\mu = \delta_{-1} - 2\delta_0 + \delta_1$, hence $|B'\mu| = \delta_{-1} + 2\delta_0 + \delta_1$ (here δ_p denotes the Dirac measure at the point p). Since $|\mu| = |f|dx$, it follows that $B'|\mu| = -D|\mu| = -\delta_{-1} + \delta_1$, hence $|B'|\mu|| = \delta_{-1} + \delta_1$. This shows that $|B'|\mu|| \neq |B'\mu|$, so B' does not satisfy the Kato equality.

Remark 3.12. a) In Example 3.11 one has $\overline{D(B')} = L^1(\mathbb{R})$, and the part A of B' in $L^1(\mathbb{R})$ generates a positive C_0-group (given by the right shift). Thus the part of B' in $\overline{D(B')}$ does satisfy (K).

b) More generally, $\overline{D(A)}$ is a band if A is a resolvent positive operator on a KB-space ([**AB**, Appendice]).

We conclude giving an example where $\pm A$ is resolvent positive, E is a reflexive Banach lattice, but neither A nor $-A$ generate C_0-semigroups.

Example 3.13. a) Let $(F, \|\ \|_F)$ be a Banach function space on $(0,\infty)$ corresponding to the function norm $\|\ \|_F$ given by

$$\|f\|_F = \|f\|_{L^p(0,\infty)} + \|f\|_{L^q(1,\infty)}$$

where $1 < p < q < \infty$. Then $(T(t)f)(x) = f(e^t x)$ defines a lattice C_0-semigroup on F. Let B be its generator. Then $\sigma(B) = \{\lambda \in \mathbb{C} : Re\ \lambda = -\frac{1}{p}\}$ and $R(\lambda, B) \geq 0$ for $\lambda > -\frac{1}{p}$, $R(\lambda, B) \leq 0$ for $\lambda < -\frac{1}{p}$. But $-B$ is not generator of a C_0-semigroup.

b) Taking $E = F \oplus F$ and $A = B \oplus (-B)$ one obtains the desired example.

Proof of a). Let $G = L^p(0, \infty)$. Then $(U(t)f)(x) = f(e^t x)$ defines a positive C_0-group on G. Let A be its generator. Then $\sigma(A) = \{\lambda \in \mathbb{C} : Re\ \lambda = -\frac{1}{p}\}$ and $(R(\lambda, A)f)(x) = \int_0^\infty e^{-\lambda t} f(e^t x)\ dt = x^\lambda \int_x^\infty f(s) s^{-\lambda - 1}\ ds$ for $\lambda > -\frac{1}{p}$.

One has $U(t)F \subset F$ and $T(t) = U(t)_{|F} \quad (t \geq 0)$. Thus B is the part of A in F.

Observe that $R(\lambda, A)G \subset F \quad (\lambda > -\frac{1}{p})$. In fact, let $0 \leq f \in G$, $g(x) = (R(\lambda, A)f)(x) = x^\lambda \int_x^\infty f(s) s^{-\lambda - 1}\ ds$. Then

$$g(x) \leq x^\lambda \|f\|_p \left(\int_x^\infty s^{(-\lambda - 1)p'}\ ds \right)^{\frac{1}{p'}} \leq \text{const } \cdot \|f\|_p \cdot x^{-\frac{1}{p}}$$

for $x \geq 1$ (where $\frac{1}{p} + \frac{1}{p'} = 1$). Thus $g \cdot 1_{(1,\infty)} \in L^\infty(1, \infty) \cap L^p(1, \infty) \subset L^q(1, \infty)$.

It follows that $(-\frac{1}{p}, \infty) \subset \varrho(B)$ and $R(\lambda, B) = R(\lambda, A)_{|F} \geq 0 \quad (\lambda > -\frac{1}{p})$.

Since, for $\lambda > -\frac{1}{p}$, $D(A) = R(\lambda, A)G \subset F$, we have $R(\lambda, A)G \subset F$ for all $\lambda \in \varrho(A)$. Thus, for $\lambda < -\frac{1}{p}$, $\lambda \in \varrho(B)$ and $R(\lambda, B) = R(\lambda, A)_{|F} \leq 0$.

Assume that $-B$ generates a C_0-semigroup $(T(-t))_{t \geq 0}$. Then $T(t)f = \lim\limits_{n \to \infty} (I + \frac{t}{n} B)^{-n} f$ in F for all $f \in F$. Since F is continuously embedded into G, it follows that $T(t) = U(-t)$.

However, $U(-t)F \not\subset F$, $t > 0$, which is a contradiction. In fact, let $-\frac{1}{p} < \alpha < -\frac{1}{q}$ and $f(x) = (1 - x)^\alpha 1_{(0,1)}(x)$. Then

$$\int_0^\infty |f(x)|^p\ dx = \int_0^1 (1 - x)^{\alpha p}\ dx = \int_0^1 y^{\alpha p}\ dy = \frac{1}{\alpha p + 1} < \infty.$$

Thus $f \in F$. However, for $t > 0$, $U(-t)f \notin L^q(1, \infty)$. In fact,

$$\|U(-t)f\|^q_{L^q(1,\infty)} = \int_1^\infty f(e^{-t} x)^q\ dx = \int_{e^{-t}}^1 (1 - y)^{\alpha q}\ dy\ e^t$$

$$= \int_0^{1 - e^{-t}} y^{\alpha q}\ dy\ e^t = \infty$$

since $\alpha q + 1 < 0$. □

In Section 4 a whole class of operators A is given for which $\pm A$ is resolvent positive but $D(A)$ is not dense. The preceding example has the additional remarkable property that the semigroup operators $T(t)$ which always exist as operators from $D(A^2)$ into E are not bounded on E. Recall, that A

generates a twice integrated semigroup S and $T(t)x = \frac{d^2}{dt^2}S(t)x$ exists for all $x \in D(A^2)$.

4. The Cesaro operator on arbitrary interpolation spaces.

In this section we shall illustrate how the theory developed in the previous section can be used to obtain results analogous to the ones in Section 2, but now for a much larger class of function spaces.

Let E be an exact (L^1, L^∞)-interpolation space on $(0, \infty)$. As in Section 2 we denote by $\mathcal{T}_E = \{T(t)\}_{t\in\mathbb{R}}$ the group defined by $T(t)f(x) = f(e^{-t}x)$. Since we do not assume that E has order continuous norm, the group $\mathcal{T}_E$ need not be strongly continuous (see Theorem 2.4). For $t \geq 0$ define

$$S_+(t)f(x) = \int_0^t T(s)f(x)\ ds = \int_0^t f(e^{-s}x)\ ds\ ,\ x > 0\ ,\ f \in E.$$

Using that E is an (L^1, L^∞)-interpolation space, it follows that $S_+(t)$ is a bounded linear operator in E and $\|S_+(t)\|_E \leq e^t - 1$ for all $t \geq 0$. Moreover, $\|S_+(t+h) - S_+(t)\|_E \leq he^{t+h}$ for all $t, h \geq 0$. Similarly, if we define

$$S_-(t)f(x) = \int_0^t T(-s)f(x)\ ds = \int_0^t f(e^{s}x)\ ds,\ x > 0,\ f \in E,$$

then $\|S_-(t)\|_E \leq \max(1 - e^{-t}, t)$ and $\|S_-(t+h) - S_-(t)\|_E \leq he^{-t}$ for all $t, h \geq 0$. We show next that $\{S_+(t)\}_{t\geq 0}$ and $\{S_-(t)\}_{t\geq 0}$ are actually integrated semigroups in E. To this end, for $\lambda \in \mathbb{C}$ with $Re\ \lambda > 1$ define

$$R(\lambda)f(x) = x^{-\lambda}\int_0^x u^{\lambda-1}f(u)du,\ x > 0,\ f \in E.$$

Via interpolation, $R(\lambda)$ is a bounded linear operator in E and

$$\|R(\lambda)\|_E \leq (Re\,\lambda - 1)^{-1}. \tag{4.1}$$

Similarly, for $Re\,\lambda < 0$ we define

$$R(\lambda)f(x) = -x^{-\lambda}\int_x^\infty u^{\lambda-1}f(u)du,\ x > 0,\ f \in E;$$

then

$$\|R(\lambda)\|_E \leq (-Re\,\lambda)^{-1}. \tag{4.2}$$

Now it is not difficult to verify that $R(\lambda) = R(\lambda, A_E)$ on $\{Re\,\lambda > 1\} \cup \{Re\,\lambda < 0\}$, where $A_E : D(A_E) \to E$ is given by

$$D(A_E) = \{f \in E : f \in AC_{\text{loc}}(0, \infty),\ xf'(x) \in E\},\ A_E f(x) = -xf'(x).$$

Moreover, integration by parts shows that

$$R(\lambda, A_E)f = \lambda \int_0^\infty e^{-\lambda t} S_+(t) f dt, \ \ Re\,\lambda > 1, \ \ f \in E$$

and

$$R(\lambda, -A_E)f = \lambda \int_0^\infty e^{-\lambda t} S_-(t) f dt, \ \ Re\,\lambda > 0, \ \ f \in E.$$

Hence, $\{S_+(t)\}_{t\geq 0}$ and $\{S_-(t)\}_{t\geq 0}$ are the integrated semigroups generated by A_E and $-A_E$ respectively. In particular, $\pm A_E$ are resolvent positive.

Remark 4.1. From the estimates (4.1) and (4.2) on $R(\lambda, A_E)$ above, it follows that the part of A_E in $\overline{D(A_E)}$ generates a strongly continuous group (cf. [**A1**, Corollary 4.2]). It is easy to see that this group is the restriction of $\mathcal{T}_E$ to $\overline{D(A_E)}$. This implies that $\overline{D(A_E)} = \{f \in E : \lim_{t\to 0} \|T(t)f - f\|_E = 0\}$. In combination with Theorem 2.4, this shows that $D(A_E)$ is dense if and only if E has order continuous norm.

Theorem 4.2. *Let E and A_E be as above. Then*

$$\sigma(A_E) = \{\lambda \in \mathbb{C} : \underline{\alpha}_E \leq Re\,\lambda \leq \overline{\alpha}_E\},$$

where $\underline{\alpha}_E$ and $\overline{\alpha}_E$ denote the lower- and upper-Boyd indices of E.

Proof. We divide the proof in five steps.

(1) As before, we denote $s^+(A_E) = \sup\{Re\,\lambda : \lambda \in \sigma(A_E)\}$ and $s^-(A_E) = s^+(-A_E)$. Then $\sigma(A_E) \subseteq \{\lambda \in \mathbb{C} : -s^-(A_E) \leq Re\,\lambda \leq s^+(A_E)\}$. Moreover, $s^+(A_E), \ -s^-(A_E) \in \sigma(A_E)$ as $\pm A_E$ are resolvent positive.

(2) Next we may use Corollary 3.8 to conclude that

$$\sigma(A_E) \cap \mathbb{R} = [-s^-(A_E), \ s^+(A_E)].$$

Indeed, from the explicit form of A_E given above it follows immediately that A_E satisfies the Kato equality and hence (K_0). Furthermore, using the representation of $R(\lambda, A_E)$ as an integral operator for $Re\,\lambda > 1$ it is easily seen that $R(\lambda, A_E)$ has no nontrivial reducing bands.

(3) For $\tau \in \mathbb{R}$ we define the isometry M_τ in E by $M_\tau f(x) = x^{i\tau} f(x), \ x > 0$. Then $M_\tau^{-1} A_E M_\tau = A_E - i\tau$, and hence $\sigma(A_E) = \sigma(A_E) - i\tau$ for all $\tau \in \mathbb{R}$.

A combination of (1), (2) and (3) already shows that

$$\sigma(A_E) = \{\lambda \in \mathbb{C} : -s^-(A_E) \leq Re\,\lambda \leq s^+(A_E)\}.$$

(4) We will show now that $s^+(A_E) = \overline{\alpha}_E$. Take $\omega > \overline{\alpha}_E$. From the definition of $\overline{\alpha}_E$ it follows that there exists $M_\omega > 0$ such that $\|D_s\|_E \le M_\omega s^\omega$ for all $s \ge 1$. Since

$$S_+(t)f(x) = \int_0^t f(e^{-s}x)\ ds = \int_1^{e^t} D_u f(x)\frac{du}{u}$$

for all $f \in E,\ x > 0$ and $t \ge 0$, it follows from Lemma 1.1 that

$$\|S_+(t)\|_E \le \int_1^{e^t} \|D_u\|_E \frac{du}{u} \le \frac{M_\omega}{\omega} e^{\omega t}$$

for all $t \ge 0$. Hence, if $Re\,\lambda > w$ then the integral $\int_0^\infty e^{-\lambda t} S_+(t)\ dt$ is convergent and $R(\lambda) = \lambda \int_0^\infty e^{-\lambda t} S_+(t)\ dt$ is analytic on $\{Re\,\lambda > \omega\}$. Therefore $s^+(A_E) \le \omega$, and this shows that $s^+(A_E) \le \overline{\alpha}_E$. Now take $\lambda > s^+(A_E)$. For $t \ge 0$ we have (since $\lambda > 0$) :

$$R(\lambda, A_E) = \int_0^\infty e^{-\lambda s} dS_+(s) \ge \int_0^t e^{-\lambda s} dS_+(s) \ge e^{-\lambda t} S_+(t),$$

so $\|S_+(t)\|_E \le e^{\lambda t} \|R(\lambda, A_E)\|_E$. For $f \in E$ the function $s \mapsto T(s)f^*$ is increasing on $[0, \infty)$, hence

$$0 \le T(t)f^*(x) \le \int_t^{t+1} T(s)f^*(x)\ ds \le S_+(t+1)f^*(x),\ x > 0,$$

and so

$$\|T(t)f\|_E = \|T(t)f^*\|_E \le \|S_+(t+1)f^*\|_E \le (e^\lambda \|R(\lambda, A_E)\|_E) e^{\lambda t} \|f\|_E.$$

This shows that $\|T(t)\|_E \le C_\lambda e^{\lambda t}$ for all $t \ge 0$, which implies (see the beginning of Section 2) that $\overline{\alpha}_E \le \lambda$. Hence $\overline{\alpha}_E \le s^+(A_E)$.

(5) Finally we show that $s^-(A_E) = -\underline{\alpha}_E$. To prove that $s^-(A_E) \le -\underline{\alpha}_E$ we may assume that $\underline{\alpha}_E > 0$, as $s^-(A_E) \le 0$. Take $-\underline{\alpha}_E < \omega < 0$. From the definition of $\underline{\alpha}_E$ it follows that there exists $M_\omega > 0$ such that $\|D_s\|_E \le M_\omega s^{-\omega}$ for all $0 < s \le 1$. Via Lemma 1.1 we see that

$$S_-(\infty)f(x) := \int_0^\infty f(e^s x)\ ds = \int_0^1 D_u f(x)\frac{du}{u},\ f \in E,\ x > 0,$$

defines a bounded linear operator in E with $\|S_-(\infty)\|_E \le (-\omega)^{-1} M_\omega$. Moreover, using Lemma 1.1 again, $\|S_-(t) - S_-(\infty)\|_E \le (-\omega)^{-1} M_\omega e^{\omega t}$ for all $t \ge 0$. Hence (cf. [**A2**, Proposition 5.5]; [**HP**, Theorem 6.2.1]) $\lambda \mapsto \int_0^\infty e^{-\lambda t} dS_-(t)$ is analytic on $\{Re\ \lambda > \omega\}$ and so $s^-(A_E) \le \omega$. This shows that $s^-(A_E) \le -\underline{\alpha}_E$.

Now we show that $-\underline{\alpha}_E \leq s^-(A_E)$. We may assume that $s^-(A_E) < 0$, as $\underline{\alpha}_E \geq 0$. Take $s^-(A_E) < \lambda < 0$. Then

$$R(\lambda, -A_E) = \int_0^\infty e^{-\lambda s} dS_-(s) \geq \int_{t-1}^t e^{-\lambda s} dS_-(s) \geq e^{-\lambda(t-1)}\{S_-(t) - S_-(t-1)\}$$

for all $t \geq 1$. For $f \in E$ the function $s \mapsto T(-s)f^*$ is decreasing on $[0, \infty)$, so

$$0 \leq T(-t)f^*(x) \leq \int_{t-1}^t T(-s)f^*(x)\ ds = S_-(t)f^*(x) - S_+(t)f^*(x),\ x > 0,$$

and hence

$$\|T(-t)f\|_E = \|T(-t)f^*\|_E \leq e^{\lambda(t-1)}\|R(\lambda, -A_E)\|_E \|f\|_E$$

for all $t \geq 1$. From this estimate it follows immediately that $-\underline{\alpha}_E \leq \lambda$, and we may conclude that $-\underline{\alpha}_E \leq s^-(A_E)$. This completes the proof of the theorem. □

Corollary 4.3. *Let E be an exact (L^1, L^∞)-interpolation space on $(0, \infty)$. Then the Cesaro operator C_E is bounded on E if and only if $\overline{\alpha}_E < 1$. In that case*

$$\sigma(C_E) = \left\{\lambda \in \mathbb{C} : 1 - \overline{\alpha}_E \leq Re\left(\frac{1}{\lambda}\right) \leq 1 - \underline{\alpha}_E\right\} \cup \{0\}. \tag{4.3}$$

Proof. If $\overline{\alpha}_E < 1$ then, by the above theorem, $1 \in \varrho(A_E)$ and integration by parts gives

$$\begin{aligned} R(1, A_E)f(x) &= \int_0^\infty e^{-t} S_+(t)f(x)\ dt = \int_0^\infty e^{-t} f(e^{-t}x)\ dt \\ &= \frac{1}{x}\int_0^x f(u)du\ , a.e.\ x \in (0, \infty) \end{aligned}$$

for all $f \in E$, i.e., $R(1, A_E) = C_E$. The indentity (4.3) now follows from a combination of Theorem 4.2 with the spectral mapping theorem for resolvents. Conversely, assume that the Cesaro operator is bounded on E. It is easy to see that the integrated semigroup generated by $A - I$ is given by

$$(W(t)f)(x) = \int_0^t e^{-s} f(e^{-s}x)\ ds.$$

Since

$$(W(t)f)(x) \le \frac{1}{x}\int_0^x f(u)du \qquad (x - a.e.)$$

it follows that $\|W(t)\| \le \|C_E\|$ for all $t \ge 0$. By Lemma 1.2 this implies that $s(A - I) < 0$. □

Remark 4.4. If we assume that $\underline{\alpha}_E > 0$ then it follows by an argument similar to the above that the operator $\widetilde{C}_E$, defined by

$$\widetilde{C}_E f(x) = \int_x^\infty f(u)\frac{du}{u}, \ a.e.\ x \in (0,\infty), \ f \in E,$$

is bounded on E and

$$\sigma(\widetilde{C}_E) = \left\{\lambda \in \mathbb{C} : \underline{\alpha}_E \le Re\ \left(\frac{1}{\lambda}\right) \le \overline{\alpha}_E\right\} \cup \{0\}.$$

Indeed, if $\underline{\alpha}_E > 0$ then $0 \in \sigma(A_E)$ and $\widetilde{C}_E = -R(0, A_E)$. It should be observed that in this general situation (i.e., without any additional assumption on the norm of E) it seems that this last result cannot be obtained via a duality argument from Corollary 4.3.

As before, let E be an exact (L^1, L^∞)-interpolation space on $(0,\infty)$, and we denote by $\{T_E(t)\}_{t\in\mathbb{R}}$ the group of bounded operators in E given by $T_E(t)f(x) = f(e^{-t}x)$ for all $f \in E$. As we have seen, if E does not have order continuous norm, this group is not strongly continuous. However, there is always a natural (locally convex) topology in E with respect to which the group is continuous. For this purpose, let E'_n denote the Köthe dual (or associate space) of E, i.e.,

$$E'_n = \left\{ g \in L^0(0,\infty) : \int_0^\infty |fg|dx < \infty \quad \forall\ f \in E \right\}.$$

Every $g \in E'_n$ defines a bounded (order continuous) linear functional φ_g on E, given by $\langle f, \varphi_g\rangle = \int_0^\infty fg dx$ for all $f \in E$. In this way we can identify E'_n with subspace of the norm dual E' (and under the present assumptions on E, this subspace is norming for E). As is known, equipped with the norm $\|g\|_{E'_n} = \|\varphi_g\|_{E'_n}$, the space E'_n is an exact (L^1, L^∞)-interpolation space on $(0,\infty)$.

Proposition 4.5. *The group $\{T_E(t)\}_{t\in\mathbb{R}}$ is continuous with respect to $\sigma(E, E'_n)$, i.e., for every $f \in E$ and $g \in E'_n$ the function $t \mapsto \int_0^\infty T(t)f(x)g(x)dx$ is continuous.*

Proof. First we assume that E is satisfies the additional condition

$$(*) \qquad f^*(x) \to 0 \ \text{ as } \ x \to \infty \ \text{ for all } \ f \in E.$$

For every $g \in E_n'$ we define the seminorm p_g on E by $p_g(f) = \int_0^\infty f^*(x)g^*(x)dx$ for all $f \in E$. Note that subadditivity of p_g follows from [**BS**, Proposition 3.6 and (3.10) on p. 54]. Actually we will show that $\{T_E(t)\}_{t\in\mathbb{R}}$ is continuous with respect to the topology σ^* generated by the seminorms $\{p_g : g \in E_n'\}$. Since, by the Hardy-Littlewood quality,

$$\left|\int_0^\infty f(x)g(x)dx\right| \le \int_0^\infty f^*(x)g^*(x)dx \quad \forall\ f \in E,\ g \in E_n',$$

the result of the proposition then follows immediately.

We denote by $\mathcal{S}$ the linear span of all characteristic functions $1_{(a,b]}$ with $0 \le a < b < \infty$. We claim that $\mathcal{S}$ is dense in E with respect to σ^*. Let A be a measurable subset of $(0,\infty)$ such that $A \subseteq (0, R]$ for some $0 < R < \infty$. Then there exists a sequence $\{B_n\}_{n=1}^\infty$ of subsets of $(0, R]$, each B_n being a finite union of intervals, such that $m(A \triangle B_n) \to 0 \quad (n \to \infty)$. This implies that $(1_A - 1_{B_n})^* \to 0$ on $(0,\infty)$ as $n \to \infty$, and so, by dominated convergence, $p_g(1_A - 1_{B_n}) \to 0 \quad (n \to \infty)$ for all $g \in E_n'$. Hence $1_A \in \overline{\mathcal{S}}^{\sigma^*}$. Now take $0 \le f \in E$. Then there exists a sequence $\{f_n\}_{n=1}^\infty$ of simple functions on bounded measurable sets such that $0 \le f_n \uparrow f$ a.e. on $(0,\infty)$. Since $f^*(x) \to 0$ as $x \to \infty$, it follows that $(f - f_n)^* \downarrow 0$ on $(0,\infty)$. Hence $p_g(f - f_n) \to 0 \quad (n \to \infty)$ for all $g \in E_n'$ by dominated convergence. From this we may conclude that $f \in \overline{\mathcal{S}}^{\sigma^*}$, by which the claim is proved.

Now we show that $p_g(T_E(t)f - f) \to 0 \quad (t \to 0)$ for all $f \in E,\ g \in E_n'$. This is easily verified for $f \in \mathcal{S}$. Take $f \in E$ arbitrary, $h \in \mathcal{S}$ and $g \in E_n'$. Then

$$p_g(T_E(t)f - f) \le p_g(T_E(t)(f - h)) + p_g(T_E(t)h - h) + p_g(f - h).$$

For $-1 \le t \le 1$ we have

$$\begin{aligned} p_g(T_E(t)(f-h)) &= \int_0^\infty [T_E(t)(f-h)]^* g^* dx \\ &= \int_0^\infty T_E(t)(f-h)^* g^* dx \le \int_0^\infty (f-h)^*(e^{-1}x)g^*(x)dx \\ &= \int_0^\infty (f-h)^*(x)g^*(ex)e\,dx = p_{g_1}(f-h), \end{aligned}$$

where $g_1 \in E_n'$ is given by $g_1(x) = eg^*(ex)$. This shows that

$$\limsup_{t\to 0} p_g(T_E(t)f - f) \leq p_{g_1}(f-h) + p_g(f-h)$$

for all $h \in \mathcal{S}$. Since $\overline{\mathcal{S}}^{\sigma *} = E$, we may conclude that $\lim_{t\to 0} p_g(T_E(t)f - f) = 0$. Observe that for $f \in E,\ g \in E_n'$ and $s \in \mathbb{R}$ we have

$$\begin{aligned} p_g(T_E(s)f) &= \int_0^\infty T_E(s)f^* \cdot g^* dx = e^s \int_0^\infty f^*(x) g^*(e^s x) dx \\ &= e^s \int_0^\infty f^* [T_{E_n'}(-s)g]^* dx = p_{g_s}(f), \end{aligned}$$

where $g_s = e^s T_{E_n'}(-s)g$. From this it follows that

$$\lim_{t\to s} p_g(T(t)f - T(s)f) = 0$$

for all $f \in E,\ g \in E_n'$ and $s \in \mathbb{R}$. This concludes the proof of the proposition in the case that E satisfies $(*)$.

Now assume that E does not satisfy $(*)$. Then $1 \in E$ and so $E_n' \subseteq L^1$, which implies that E_n' satisfies $(*)$. Since E is a subspace of $(E_n')_n'$, it follows from the first part of the proof that

$$\lim_{t\to 0} \int_0^\infty f \cdot T_{E_n'}(t) g dx = \int_0^\infty f g dx$$

for all $f \in E$ and $g \in E_n'$. Since

$$\int_0^\infty T_E(t) f \cdot g dx = e^t \int_0^\infty f T_{E_n'}(-t) g dx,$$

this implies that $\lim_{t\to s} \int_0^\infty T_E(t) f \cdot g dx = \int_0^\infty f g dx$ for all $f \in E$ and $g \in E_n'$. This suffices to prove the proposition in this case.

5. The Black-Scholes partial differential equation in (L^1, L^∞)-interpolation spaces.

The Black-Scholes partial differential equation is a degenerate parabolic equation of the form

$$u_t = x^2 u_{xx} + x u_x \qquad (t > 0,\ x > 0). \tag{5.1}$$

The aim of this section is to discuss its well-posedness, spectral properties and asymptotic behaviour in (L^1, L^∞)-interpolation spaces. It is convenient to consider the corresponding operator

$$\begin{aligned} B : \mathcal{D}(0,\infty)' &\to \mathcal{D}(0,\infty)' \\ Bf &= x^2 f'' + x f'; \\ \text{i.e., } \langle Bf, \varphi \rangle &= \langle f, ((m^2\varphi)' - m\varphi)' \rangle \end{aligned}$$

for all $\varphi \in \mathcal{D}(0,\infty)$, $f \in \mathcal{D}(0,\infty)'$ where $m(x) = x \qquad (x > 0)$.

Given an (L^1, L^∞)-interpolation space E we consider the part B_E of B in E; i.e., B_E is the operator on E with domain

$$\begin{aligned} D(B_E) &= \{ f \in E : Bf \in E \} \\ B_E f &= Bf \ . \end{aligned}$$

Here we use that $E \subset L^1_{\mathrm{loc}}(0,\infty) \subset \mathcal{D}(0,\infty)'$ with the usual identification of functions with distributions. The following proposition allows us to use the results of the preceding sections.

Proposition 5.1. *Let E be an (L^1, L^∞)-interpolation space. Then $B_E = (A_E)^2$.*

Proof. Recall that $D(A_E) = \{ f \in E : mf' \in E \}$, $A_E f = -mf'$.

a) We show that $\lambda^2 - B_E$ is injective for $\lambda > 1$. Let $k \in D(B_E)$ such that $(\lambda^2 - B_E)k = 0$. Let $h = \lambda k + mk' \in \mathcal{D}(0,\infty)'$. Then $\lambda h - mh' = 0$ in $\mathcal{D}(0,\infty)'$. This implies that $h \in C(0,\infty)$ and

$$(x^{-\lambda} h)' = x^{-\lambda - 1}(-\lambda h + x h') = 0.$$

Hence $h(x) = c x^\lambda$ for some constant c. Thus $\lambda k(x) + x k'(x) = c x^\lambda \in \mathcal{D}(0,\infty)'$. Hence $k \in C^\infty(0,\infty)$ and

$$(x^\lambda k)' = x^{\lambda - 1}(\lambda k + x k') = c x^{2\lambda - 1}.$$

This implies that $x^\lambda k = a x^{2\lambda} + b$ for some constants a and b. We have shown that $k(x) = a x^\lambda + b x^{-\lambda}$ which is in $L^1 + L^\infty$ only if $a = b = 0$.

b) Now let $f \in D(B_E)$. Let $\lambda > 1$. Then $\lambda \in \varrho(\pm A_E)$. Hence $\lambda^2 \in \varrho(A_E^2)$ and $R(\lambda^2, A_E^2) = (\lambda - A_E)^{-1}(\lambda + A_E)^{-1}$. Let $k = R(\lambda^2, A_E^2)(\lambda^2 - B_E)f$. Then $k \in D(A_E^2)$. Since A_E^2 is a restriction of B_E we have $(\lambda^2 - B_E)k = (\lambda^2 - B_E)f$. Since $(\lambda^2 - B_E)$ is injective, it follows that $f = k \in D(A_E^2)$. $\square$

As a first consequence we determine the spectrum of B_E.

Theorem 5.2. *Let E be an (L^1, L^∞)-interpolation space with Boyd indices $\underline{\alpha}_E$ and $\overline{\alpha}_E$. Then*

$$\sigma(B_E) = \left\{ r + is : \underline{\alpha}_E^2 - \frac{s^2}{4\underline{\alpha}_E^2} \le r \le \overline{\alpha}_E^2 - \frac{s^2}{4\overline{\alpha}_E^2} \right\};$$

i.e., $\sigma(B_E)$ is the region between two parabolas (with appropriate modification if $\underline{\alpha}_E = 0$ or $\underline{\alpha}_E = \overline{\alpha}_E = 0$).

Proof. By Theorem 4.2 we have

$$\sigma(A_E) = \{\lambda \in \mathbb{C} : \underline{\alpha}_E \le Re\ \lambda \le \overline{\alpha}_E\}.$$

Since $\sigma(B_E) = \sigma(A_E)^2$ it follows that

$$\begin{aligned} \sigma(B_E) &= \{\alpha^2 + 2\alpha\beta i - \beta^2 : \beta \in \mathbb{R},\ \underline{\alpha}_E \le \alpha \le \overline{\alpha}_E\} \\ &= \left\{\alpha^2 - \frac{s^2}{4\alpha^2} + is : s \in \mathbb{R},\ \underline{\alpha}_E \le \alpha \le \overline{\alpha}_E\right\} \end{aligned}$$

which implies the claim. □

Thus the spectrum of B_E varies very much as a function of the (L^1, L^∞)-interpolation space.

Next we consider the semigroup generated by B_E.

Theorem 5.3. *Let E be an (L^1, L^∞)-interpolation space with order continuous norm. Then B_E generates a holomorphic C_0-semigroup V_E on E of angle $\pi/2$. Moreover, the exponential type $\omega(V_E)$ of V_E is given by*

$$\omega(V_E) = (\overline{\alpha}_E)^2.$$

Proof. This follows directly from the fact that $B_E = (A_E)^2$ and that A_E generates a C_0-group (cf. [**N**, Theorem 1.15]). It follows from Theorem 5.2 that $s(B_E) = (\overline{\alpha}_E)^2$. Since V_E is holomorphic, $s(B_E) = \omega(V_E)$. □

If E does not have order continuous norm, then $D(B_E)$ is not dense. Still the holomorphic estimate for the resolvent is valid. This situation is very well studied by E. Sinestrari [**Si**] from which we quote the following result.

Theorem 5.4. *Let A be an operator on a Banach space X. Assume that there exist $w \in \mathbb{R}$, $\theta \in [0, \pi/2]$ such that*

$$\begin{cases} w + \Sigma(\theta + \pi/2) \subset \varrho(A) \ \textit{and} \\ \|\lambda R(\lambda, A)\| \le M \ \textit{if}\ \lambda \in w + \Sigma(\theta + \pi/2). \end{cases} \tag{5.2}$$

Then there exists a holomorphic mapping

$$T : \Sigma(\theta) \to \mathcal{L}(X)$$

such that $T(z + z') = T(z)T(z') \qquad (z, z' \in \Sigma(\theta))$,

$$\sup_{|Arg z| < \theta'} \|e^{-wz}T(z)\| < \infty \ \textit{for all}\ 0 < \theta' < \theta, \tag{5.3}$$

and

$$R(\lambda, A) = \int_0^\infty e^{-\lambda t} T(t)\ dt \qquad (\lambda > \omega). \tag{5.4}$$

In that case, we call T the **generalized holomorphic semigroup** *generated by A.*

Here we used the usual notation

$$\Sigma(\theta) = \{re^{i\alpha} : r > 0,\ \alpha \in (-\theta, \theta)\}.$$

The semigroup T has the following regularity property. Considering $D(A^k)$ as a Banach space for the norm $\|x\|_{D(A^k)} = \|x\| + \|Ax\| + \ldots + \|A^k x\|$, one has

$$T(\cdot)x \in C^\infty((0,\infty),\ D(A^k)) \text{ and} \tag{5.5}$$

$$\frac{d}{dt}T(t)x = AT(t)x \qquad (t > 0) \tag{5.6}$$

for all $x \in X,\ k \in \mathbb{N}$, see [**Si**] for this. Denoting by

$$s(A) = \sup\{Re\,\lambda : \lambda \in \sigma(A)\}$$

the spectral bound of A, as before, and by

$$\omega(T) = \inf\left\{w \in \mathbb{R} : \sup_{t\geq 0} \|e^{-\omega t}T(t)\| < \infty\right\}$$

the type of T, one has as in the strongly continuous case

$$s(A) = \omega(T). \tag{5.7}$$

Proof of (5.7). Let $Y = \overline{D(A)} \subset X$ and denote by A_0 the part of A in Y. Then A_0 generates a holomorphic C_0-semigroup $(T_0(t))_{t\geq 0}$ on Y and one has $T_0(t) = T(t)_{|Y}$. Since $D(A) \subset Y$ one has $\sigma(A) = \sigma(A_0)$ (by [**A4**, Proposition 1.1]), and in particular $s(A) = s(A_0)$. Let $w' > s(A)$. Then

$$\|T_0(t)\|_{\mathcal{L}(Y)} \leq M' e^{w' t} \qquad (t \geq 0).$$

Since $T(1)X \subset Y$, it follows that $\omega(T) \leq \omega'$. □

Now we can formulate the following result for the operator B_E.

Proposition 5.5. *Let E be an (L^1, L^∞)-interpolation space. Then B_E generates a generalized holomorphic semigroup V_E on E. The semigroup V_E is strongly continuous if and only if E has order continuous norm. Finally the exponential type of V_E is given by*

$$\omega(V_E) = (\overline{\alpha}_E)^2. \tag{5.8}$$

Proof. It follows from (4.1) and (4.2) that

$$\|R(\lambda, A_E)\| \leq (|Re\,\lambda| - 1)^{-1} \qquad (|Re\,\lambda| \geq 1). \tag{5.9}$$

Now the argument given in [**N**] A-II Theorem 1.14 and 1.15 shows that $B_E = A_E^2$ satisfies (5.2). Hence B_E generates a generalized holomorphic semigroup.

If $D(B_E)$ is dense, then also $D(A_E)$ is dense, since $D(B_E) \subset D(A_E)$. Conversely, assume that $D(A_E)$ is dense. Then $\lambda R(\lambda, A_E) \to I$ strongly as $\lambda \to \infty$. Hence $(\lambda R(\lambda, A_E))^2 \to I$ strongly as $\lambda \to \infty$. Thus $D(B_E) =$

$D(A_E^2)$ is dense. Now the second claim follows from Theorem 2.4. Finally, Theorem 5.2 and (5.7) imply that $\omega(V_E) = s(B_E) = (\overline{\alpha}_E)^2$. □

Next we establish the usual formula for V_E.

Proposition 5.6. *Let E be an (L^1, L^∞)-interpolation space. Then*

$$\langle V_E(t)f,\ \varphi\rangle = (4\pi t)^{-1/2} \int_{\mathbb{R}} e^{-r^2/4t} \langle T_E(r)f, \varphi\rangle dr \tag{5.10}$$

for all $f \in E,\ \varphi \in E'$.

Proof. We use the following formula

$$\frac{e^{-\lambda|r|}}{2\lambda} = \int_0^\infty e^{-\lambda^2 t} (4\pi t)^{-1/2} e^{-r^2/4t}\ dt, \tag{5.11}$$

valid for all $\lambda > 0,\ r \in \mathbb{R}$ (see [**D**, p. 138]).

For $\lambda > 1$ we have

$$\begin{aligned}
\int_0^\infty e^{-\lambda^2 t} V_E(t)\ dt &= R(\lambda^2, A_E^2) \\
&= (\lambda - A_E)^{-1}(\lambda + A_E)^{-1} = -R(\lambda, A_E)R(-\lambda, A_E) \\
&= \frac{R(\lambda, A_E) + R(-\lambda, A_E)}{2\lambda} \\
&= \frac{1}{2\lambda}\left(\int_0^\infty e^{-\lambda t} T_E(t)\ dt + \int_0^\infty e^{-\lambda t} T_E(t)\ dt\right) \\
&= \int_{-\infty}^{+\infty} \frac{e^{-\lambda|r|}}{\lambda} T_E(r)\ dr \\
&= \int_{-\infty}^{+\infty} T_E(r) \int_0^\infty e^{-\lambda^2 t}(4\pi t)^{-1/2} e^{-r^2/4t}\ dt\ dr \\
&= \int_0^\infty e^{-\lambda^2 t} \int_{\mathbb{R}} (4\pi t)^{-1/2} e^{-r^2/4t} T_E(r)\ dr\ dt.
\end{aligned}$$

Here the integrals involving $T_E(t)$ are understood in the $\sigma(E, E')$-duality. Observe that it suffices to evaluate by $f \in E_+$ and $\varphi \in E'_+$ only, so that Fubini's theorem can be applied. Now the claim follows from the uniqueness theorem for Laplace transforms. □

It is easy to deduce a pointwise expression from (5.10):

$$(5.12)\qquad (V_E(t)f)(x) = (4\pi t)^{-1/2}\int_{\mathbb{R}} e^{-r^2/4t} f(e^{-r}x)\,dr = (4\pi t)^{-1/2}\int_0^\infty e^{-(\log x-\log y)^2/4t} f(y)\frac{dy}{y}.$$

Thus V_E is an integral operator.

From Proposition 5.6 we now deduce the following continuity result.

Proposition 5.7. *Let E be an exact interpolation space and E'_n its Köthe dual. Then V_E is $\sigma(E,E'_n)$-continuous, i.e.,*

$$\lim_{t\downarrow 0}\langle V_E(t)f,\varphi\rangle = \langle f,\varphi\rangle$$

for all $f\in E,\ \varphi\in E'_n$.

Proof. Let $f\in E,\ \varphi\in E'_n$. Let $\varepsilon>0$. By Proposition 4.5 we can choose $\delta>0$ such that $|\langle T_E(r)f,\varphi\rangle-\langle f,\varphi\rangle|\le\varepsilon$ if $|r|\le\delta$. Then

$$\begin{aligned}
&\limsup_{t\downarrow 0}|\langle V_E(t)f,\varphi\rangle-\langle f,\varphi\rangle|\\
&=\limsup_{t\downarrow 0}(4\pi t)^{-1/2}\left|\int_{\mathbb{R}} e^{-r^2/4t}(\langle T_E(r)f,\varphi\rangle-\langle f,\varphi\rangle)dr\right|\\
&\le \varepsilon+\limsup_{t\downarrow 0}(4\pi t)^{-1/2}\int_{|r|\ge\delta} e^{-r^2/4t}|\langle T_E(r)f,\varphi\rangle-\langle f,\varphi\rangle|dr\\
&=\varepsilon.
\end{aligned}$$

This implies the claim. □

Now we obtain the following final result on existence and uniqueness for the Black & Scholes partial differential equation.

Theorem 5.8. *Let E be an exact (L^1,L^∞)-interpolation space with Köthe dual E'_n. Let $f\in E,\ u(t)=V_E(t)f$. Then u is the unique solution of the Cauchy problem*

$$(\mathrm{CP})\qquad \begin{cases} u\in C^1((0,\infty);E), & u(t)\in D(B_E)\quad (t>0);\\ \dot u(t)=B_E u(t) & (t>0)\\ \lim_{t\downarrow 0} u(t)=f & \text{for } \sigma(E,E'_n).\end{cases}$$

Moreover, if we put $u(t,x)=(V_E(t)f)(x)=u(t)(x)$, then $u\in C^\infty(0,\infty)\times(0,\infty)$ and

$$(\mathrm{BS})\qquad u_t = x^2u_{xx}+xu_x \qquad (t>0,\ x>0).$$

Proof. We know that u is a solution of (CP). In order to prove uniqueness let u be a solution of (CP) with $f = 0$. Let $t > 0$, $v(s) = V_E(t-s)u(s)$, $s \in (0,t)$. Since V_E is holomorphic and $\frac{d}{dt}V_E(t) = B_E V_E(t) \quad (t > 0)$ we have

$$\dot{v}(s) = -B_E V_E(t-s)u(s) + V_E(t-s)\dot{u}(s) = 0.$$

Thus v is constant on $(0,t)$. Moreover, $V_E(t-s) \to V_E(t)$ as $s \downarrow 0$ in $\mathcal{L}(E)$. Let $\varphi \in E_n'$. Then

$$\begin{aligned} \langle v(s), \varphi \rangle &= \langle (V_E(t-s) - V_E(t))u(s), \varphi \rangle \\ &\quad + \langle u(s), V_E(t)'\varphi \rangle \\ &\to 0 \quad (s \downarrow 0). \end{aligned}$$

Here we use that $V_E(t)'\varphi \in E_n'$ which follows from (5.10). Thus $v(s) \equiv 0$ on $(0,t)$. Since $u(s) \to u(t)$ in norm as $s \uparrow t$ and $V_E(t-s)u(t) \to u(t)$ for $\sigma(E, E_n')$ as $s \uparrow t$, it follows that $v(s) = V_E(t-s)(u(s) - u(t)) + V_E(t-s)u(t) \to u(t)$ as $s \uparrow t$ for $\sigma(E, E_n')$. Thus $u(t) = 0$.

It remains to show the regularity result. For $f \in D(A_E)$ we have $f \in L^1_{\text{loc}}(0,\infty)$ and $xf' \in E \subset L^1_{\text{loc}}(0,\infty)$. Hence $f \in C(0,\infty)$. From this one obtains by induction that $D(A_E^{k+1}) \subset C^k(0,\infty)$ for all $k \in \mathbb{N}$. Now we know that $V_E(\cdot)f \in C^\infty((0,\infty), D(B_E^k)) = C^\infty((0,\infty); D(A_E^{2k}))$ for all $f \in E$. It is not difficult to see that this implies that $u \in C^\infty(0,\infty) \times (0,\infty)$. □

6. Perturbation.

Let B be an operator on a Banach space X. An operator $Q : D(B) \to X$ is called a **small perturbation** of B if for all $\varepsilon > 0$ there exists $b \geq 0$ such that

$$\|Qx\| \leq \varepsilon \|Bx\| + b\|x\| \qquad (x \in D(B)). \tag{6.1}$$

The following is well-known.

Proposition 6.1. *Let B be the generator of a (generalized) holomorphic semigroup and let Q be a small perturbation of B. Then $B + Q$ generates a (generalized) holomorphic semigroup.*

Example 6.2. Let E be an (L^1, L^∞)-interpolation space. Then A_E is a small perturbation of B_E.

Proof. We have for $\lambda > 1$,

$$R(\lambda^2, B_E) = \frac{1}{2\lambda}(R(\lambda, A_E) + R(\lambda, -A_E))$$

(see the proof of Proposition 5.6). Hence

$$\begin{aligned}\|A_E R(\lambda^2, B_E)\| &= \frac{1}{2\lambda}\|A_E R(\lambda, A_E) + A_E R(\lambda, -A_E)\| \\ &= \frac{1}{2\lambda}\|\lambda R(\lambda, A_E) - \lambda R(\lambda, -A_E)\| \\ &\leq \frac{1}{2}(\|R(\lambda, A_E)\| + \|R(\lambda, -A_E)\|) \\ &\to 0 \quad (\lambda \to \infty).\end{aligned}$$

Let $\varepsilon > 0$. Choose $\lambda > 1$ such that $\|A_E R(\lambda, B_E^2)\| \leq \varepsilon$. Let $f \in D(B_E)$. Then $\|A_E f\| = \|A_E R(\lambda^2, B_E)(\lambda^2 - B_E)f\| \leq \varepsilon\|(\lambda^2 - B_E)f\| \leq \varepsilon\|B_E f\| + \lambda^2\|f\|$. □

It remains to show that $\sigma(E, E_n')$-continuity is preserved by small perturbations. For this we establish a Tauberian theorem (Proposition 6.4) which is valid for Laplace transforms of functions having a holomorphic extension to a sector. They can be characterized as follows (see Prüß [**P**, Theorem 0.1]).

Proposition 6.3. *Let X be a Banach space and let $0 < \theta_0 \leq \pi/2$.*

a) *Let $r : \Sigma(\theta_0 + \pi/2) \to X$ be a holomorphic function such that*

$$\sup_{\lambda \in \Sigma(\theta + \pi/2)} \|\lambda r(\lambda)\| < \infty \tag{6.2}$$

for all $0 < \theta < \theta_0$. Then there exists a holomorphic function $f : \Sigma(\theta_0) \to X$ satisfying

$$\sup_{z \in \Sigma(\theta)} \|f(z)\| < \infty \tag{6.3}$$

for all $0 < \theta < \theta_0$ such that $r(\lambda) = \hat{f}(\lambda) := \int_0^\infty e^{-\lambda t} f(t)dt$ for $Re\,\lambda > 0$.

b) *Conversely, assume that $f : \mathbb{R}_+ \to X$ has a holomorphic extension to $\Sigma(\theta_0)$ satisfying* (6.3); *then the Laplace transform $\hat{f}$ of f has a holomorphic extension r to $\Sigma(\theta_0 + \pi/2)$ satisfying* (6.2).

Now we describe the asymptotic behaviour of $f(t)$ for $t \downarrow 0$ in terms of the behaviour of $r(\lambda)$ as $\lambda \to \infty$.

Proposition 6.4. *Assume that f and r are as in Proposition* 6.3. *Let $c \in X$. Then $\lim_{t\downarrow 0} f(t) = c$ if and only if $\lim_{\lambda\to\infty} \lambda r(\lambda) = c$.*

Proof. 1. Assume that $\lim_{\lambda\to\infty} \lambda r(\lambda) = c$. Choose $0 < \theta < \theta_0$. It follows from [**HP**, Theorem 3.14.3] that

$$\lim_{\substack{|\lambda|\to\infty \\ \lambda \in \Sigma(\theta+\pi/2)}} \lambda r(\lambda) = c.$$

Let $\varepsilon > 0$. Choose $\varrho_0 > 0$ such that $\|\lambda r(\lambda) - c\| \le \varepsilon$ for all $\lambda \in \overline{\Sigma}(\theta + \pi/2)$ with $|\lambda| \ge \varrho_0$. Let $t \ge 1/\varrho_0$. Choose a contour Γ consisting of the lines $\{\varrho e^{\pm i(\theta+\pi/2)} : \varrho \ge 1/t\}$ and the arc $\{1/t \cdot e^{i\alpha} : -\theta \le \alpha \le \theta\}$. Then by the proof of [**P**, Theorem 0.1],

$$f(t) = \frac{1}{2\pi i} \int_\Gamma e^{\lambda t} r(\lambda) d\lambda.$$

Since $\frac{1}{2\pi i} \int_\Gamma e^{\lambda t} \frac{d\lambda}{\lambda} = 1$,

$$\begin{aligned} \|f(t) - c\| &= \left\| \frac{1}{2\pi i} \int_\gamma e^{\lambda t} (\lambda r(\lambda) - c) \frac{d\lambda}{\lambda} \right\| \\ &\le \frac{\varepsilon}{2\pi} \left\{ 2 \int_{1/t}^{\infty} e^{-tr\cos\theta} \frac{dr}{r} + \int_{-\theta}^{\theta} e^{\cos\alpha} d\alpha \right\} \\ &\le \frac{\varepsilon}{2\pi} \left\{ 2 \cdot \frac{1}{|\cos\theta|} + \int_{-\theta}^{\theta} e^{\cos\alpha} d\alpha \right\}. \end{aligned}$$

This proves the claim.

2. The converse is a classical Abelian theorem. □

Proposition 6.5. *Let A be the generator of a generalized holomorphic semigroup T on a Banach space X and let B be a small perturbation of A. Denote by S the generalized holomorphic semigroup generated by $A + B$. Let $x \in X$, $\varphi \in X'$, such that $\lim_{t\downarrow 0} \langle T(t)x, \varphi\rangle = \langle x, \varphi\rangle$. Then $\lim_{t\downarrow 0}\langle S(t)x, \varphi\rangle = \langle x, \varphi\rangle$.*

Proof. Replacing A by $A - w$ if necessary, we can assume that A and $A+B$ satisfy (5.2) with $w = 0$. So we are in the situation of Proposition 6.4. Thus we know that $\lim_{\lambda\to\infty}\langle \lambda R(\lambda, A)x, \varphi\rangle = \langle x, \varphi\rangle$, and it suffices to show that $\lim_{\lambda\to\infty}\langle \lambda R(\lambda, A + B)x, \varphi\rangle = \langle x, \varphi\rangle$. For this it suffices to show that

$$\|\lambda R(\lambda, A + B) - \lambda R(\lambda, A)\| \to 0 \qquad (\lambda \to \infty).$$

Let $M \ge 0$ such that $\|\lambda R(\lambda, A)\| \le M$ $(\lambda > 0)$. Let $\varepsilon > 0$. There exists $b \ge 0$ such that

$$\begin{aligned} \|BR(\lambda, A)\| &\le \varepsilon\|AR(\lambda, A)\| + b\|R(\lambda, A)\| \\ &\le \varepsilon\|\lambda R(\lambda, A) - I\| + b\|R(\lambda, A)\| \\ &\le \varepsilon(M + 1) + bM/\lambda. \end{aligned}$$

Thus $\lim_{\lambda\to\infty} \|BR(\lambda, A)\| \le \varepsilon(M + 1)$. □

As a result we now know the following. Let E be an (L^1, L^∞)-interpolation space. Let Q be a small perturbation of B_E. Then $B_E + Q$ generates a generalized holomorphic semigroup on E which is $\sigma(E, E'_n)$ continuous. In particular, we obtain the following result.

Theorem 6.6. *Let E be an (L^1, L^∞)-interpolation space. Let $\alpha > 0$ be a constant, and let $\beta, \gamma \in L^\infty(0,\infty)$. Consider the operator G on E given by*

$$\begin{aligned} Gf &= \alpha x^2 f'' + \beta x f' + \gamma f \\ D(G) &= D(B_E). \end{aligned}$$

Then G generates a generalized holomorphic semigroup which is $\sigma(E, E_n')$-continuous.

Proof. By Example 6.2, the operator A_E is a small perturbation of B_E. Thus $B_E - A_E$ generates a generalized holomorphic semigroup. Since β defines a bounded multiplication operator on E, $\beta A_E + \gamma$ is a small perturbation of $\alpha(B_E - A_E)$. Note that $G = \alpha(B_E - A_E) + \beta A_E + \gamma$. □

References

[A1] W. Arendt, *Kato's equality and spectral decomposition for positive C_0-groups*, Manuscripta Math., **40** (1982), 277-298, MR 85d:47041, Zbl 543.47032.

[A2] ———, *Resolvent positive operators*, Proc. London Math. Soc., **54** (1987), 321-349, MR 88c:47074, Zbl 617.47029.

[A3] ———, *Vector-valued Laplace transforms and Cauchy problems*, Israel J. Math., **59** (1987), 327-353, MR 89a:47064, Zbl 637.44001.

[A4] ———, *Gaussian estimates and interpolation of the spectrum in L^p*, Differential and Integral Equations, **7** (1994), 1153-1168, MR 95e:47066, Zbl 827.35081.

[A5] ———, *Spectrum and growth of positive semigroups*, in "Evolution Equations", G. Ferreyra, G. Goldstein, F. Neubrander eds., Marcel-Dekker, 1994, MR 95m:47062, Zbl 834.47031.

[AB] W. Arendt and Ph. Bénilan, *Inégalités de Kato et semi-groupes sous-markoviens*, Rev. Mat. Complutense Madrid, **5** (1992), 279-308, MR 93j:47062, Zbl 769.47010.

[AG] W. Arendt and G. Greiner, *The spectral mapping theorem for one-parameter groups of positive operators on $C_0(X)$*, Semigroup Forum, **30** (1984), 297-330, MR 86h:47068, Zbl 536.47032.

[ANS] W. Arendt, F. Neubrander and U. Schlotterbeck, *Interpolation of semigroups and integrated semigroups*, Semigroup Forum, **45** (1992), 26-37, MR 93e:47044, Zbl 782.47035.

[BGV] E. Barucci, F. Gozzi and V. Vespri, *On a semigroup approach to no-arbitrage pricing theory*, in "Proceedings of the Seminar on Stochastic Analysis, Random Fields and Applications," Conference held in Ascona, 1996, MR 2000j:91036, Zbl 933.91010.

[BS] C. Bennett and R. Sharpley, *Interpolation of Operators*, Academic Press, London, 1988, MR 89e:46001, Zbl 647.46057.

[BlSc] F. Black and M. Scholes, *The pricing of options and corporate liabilities*, Journal of Political Economy, **81** (1973), 637-654.

[Bo] D.W. Boyd, *The spectrum of the Cesaro operator*, Acta Sci. Math., **29** (1968), 31-34, MR 39 #798, Zbl 157.45501.

[CH] Ph. Clémènt, H.J.A.M. Heijmans et al., *One-Parameter Semigroups*, North Holland, 1987, MR 89b:47058, Zbl 636.47051.

[CGV] F. Colombo, M. Giuli and V. Vespri, *Generation of smoothing semigroups by elliptic operators and some applications to mathematical finance*, preprint, Firenze, 1998.

[D] G. Doetsch, *Handbuch der Laplace-Transformation*, Band 1, Birkhäuser, Basel, 1971, MR 49 #9546, Zbl 242.44001.

[EN] K. Engel and R. Nagel, *One-parameter Semigroups for Linear Evolution Equations*, Springer, Berlin, 2000, MR 2000i:47075, Zbl 952.47036.

[GMV] F. Gozzi, R. Monte and V. Vespri, *Generation of analytic semigroups for degenerate elliptic operators arising in financial mathematics*, Ulmer Seminare, **2** (1997), 155-196.

[G] G. Greiner, *A spectral decomposition of strongly continuous groups of positive operators*, Quart. J. Oxford, **35** (1984), 37-47, MR 85i:47038, Zbl 565.47025.

[HP] E. Hille and R.S. Phillips, *Functional Analysis and Semigroups*, Amer. Math. Soc. Coll. Publ., **31**, Providence, 1957, MR 19,664d, Zbl 078.10004.

[KPS] S.G. Krein, Ju. Petunin and E.M. Semenov, *Interpolation of Linear Operators*, Transl. Math. Monogr., **54**, Amer. Math. Soc., Providence, 1982, MR 84j:46103, Zbl 251.46038.

[N] R. Nagel, ed., *One-Parameter Semigroups of Positive Operators*, Springer, **LN 1184** (1986), MR 88i:47022, Zbl 585.47030.

[P] J. Prüß, *Evolutionary Integral Equations and Applications*, Birkhäuser, Basel, 1993, MR 94h:45010, Zbl 784.45006.

[S] H.H. Schaefer, *Banach Lattices and Positive Operators*, Springer, Berlin, 1974, MR 54 #11023, Zbl 296.47023.

[Si] E. Sinestrari, *On the abstract Cauchy problem of parabolic type in spaces of continuous functions*, J. Math. Anal. Appl., **107** (1985), 16-66, MR 86g:34086, Zbl 589.47042.

[vN] J.M.A.M. van Neerven, *The Asymptotic Behaviour of Semigroups of Linear Operators*, Birkhäuser, Basel, 1996, MR 98d:47001, Zbl 905.47001.

Received January 27, 2000.

Universität Ulm
Angewandte Analysis
D-89069 Ulm
Germany
E-mail address: arendt@mathematik.uni-ulm.de

Department of Mathematics
Faculty ITS
Delft University of Technology
P.O. Box 5031
NL-2600 GA Delft
The Netherlands
E-mail address: B.dePagter@twi.tudelft.nl

PACIFIC JOURNAL OF MATHEMATICS
Vol. 202, No. 1, 2002

MIRANDA–PERSSON'S PROBLEM ON EXTREMAL ELLIPTIC K3 SURFACES

Enrique Artal Bartolo, Hiro-o Tokunaga, and De-Qi Zhang

In one of their early works, Miranda and Persson have classified all possible configurations of singular fibers for semistable extremal elliptic fibrations on $K3$ surfaces. They also obtained the Mordell-Weil groups in terms of the singular fibers except for 17 cases where the determination and the uniqueness of the groups were not settled. In this paper, we settle these problems completely. We also show that for all cases with 'larger' Mordell-Weil groups, this group, together with the singular fibre type, determines uniquely the fibration structure of the $K3$ surface (up to based fibre-space isomorphisms).

0. Introduction.

Let $f : X \to C$ be an elliptic surface over a smooth projective curve C with a section O, i.e., a Jacobian elliptic fibration over C. Throughout this paper, we always assume that

($*$) f has at least one singular fiber.

Let $MW(f)$ be the Mordell-Weil group of $f : X \to C$, i.e., the group of sections, O being the zero. Under the assumption ($*$), it is known that $MW(f)$ is a finitely generated Abelian group (the Mordell-Weil theorem). More precisely, if we let R be the subgroup of the Néron-Severi group $\mathrm{NS}(X)$ of X generated by O and all the irreducible components in fibers of f, then (i) $\mathrm{NS}(X)$ is torsion-free, and (ii) $MW(f) \cong \mathrm{NS}(X)/R$ (see [**S**], for instance). Note that the Shioda-Tate formula $\operatorname{rank} MW(f) = \rho(X) - \operatorname{rank} R$ easily follows from the second statement.

We call $f : X \to C$ *extremal* if

(i) the Picard number $\rho(X)$ of X is equal to $h^{1,1}$ and
(ii) $\operatorname{rank} MW(f) = 0$.

If $f : X \to C$ is extremal, then the Shioda-Tate formula implies $\operatorname{rank} R = \rho(X)$. Hence, in other words, $f : X \to C$ is extremal if and only if $\rho(X) = \operatorname{rank} R = h^{1,1}(X)$. Also, taking the isomorphism $MW(f) \cong NS(X)/R$ into account, it seems that we can say a lot about $MW(f)$ only from the data of types of singular fibers.

In [**MP1**], Miranda and Persson studied extremal rational elliptic surfaces. They gave a complete classification and proved the uniqueness of such surfaces.

Suppose that $f : X \to C$ is a semi-stable elliptic $K3$ surface, i.e., f has only I_n type singular fibers with Kodaira's notation [**Ko**]. In this case, $C = \mathbf{P}^1$, $\mathrm{NS}(X) = \mathrm{Pic}\, X$, and f is extremal if and only if f has exactly six singular fibers. For a semi-stable elliptic $K3$ surface, the configuration of singular fibers is said to be $[n_1, \dots, n_s]$ $(n_1 \leq n_2 \leq \cdots \leq n_s)$ if it has singular fibers $I_{n_1}, \dots, I_{n_s}$. In [**MP2**], Miranda and Persson gave a complete list for realizable s-tuples $[n_1, \dots, n_s]$; and their list shows that there are 112 extremal cases. In [**MP3**], they go on to study $MW(f)$ for those extremal elliptic $K3$ surfaces.

We say that $f : X \to \mathbf{P}^1$ is of *type* m if the corresponding $[n_1, n_2, \dots, n_6]$ appears as the No. m case in the table of [**MP3**]. Suppose that f is of type m. What Miranda and Persson did in [**MP3**] are that

(i) if $m \neq 2, 4, 9, 11, 13, 27, 31, 32, 35, 37, 38, 44, 48, 53, 55, 69$ and 92, $MW(f)$ is determined by the 6-tuples $[n_1, n_2, \dots, n_6]$, and
(ii) if $MW(f) \supseteq \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$, then the corresponding elliptic $K3$ surface is unique.

The main purpose of this paper is

(i) to determine $MW(f)$ for the remaining cases, and
(ii) to consider the uniqueness problem for other kinds of $MW(f)$; more precisely, this problem may be formulated as follows:

Question 0.1. Let $f_1 : X_1 \to \mathbf{P}^1$ and $f_2 : X_2 \to \mathbf{P}^1$ be semi-stable extremal elliptic $K3$ surfaces such that

(i) both X_1 and X_2 have the same configuration of singular fibers, and
(ii) their Mordell-Weil groups are isomorphic.

Then is it true that there exists an isomorphism $\varphi : X_1 \to X_2$ such that

(a) φ preserves the fibrations, and
(b) the zero section of f_1 maps to that of f_2 with φ?

Now let us state our result concerning the first problem.

Theorem 0.2. *Let $f : X \to \mathbf{P}^1$ be of type m, m being one of the* 17 *exceptional cases as above. Then we have the following table:*

m	*the* 6*-tuple*	$MW(f)$	m	*the* 6*-tuple*	$MW(f)$
2	$[1,1,1,1,2,18]$	(0), $\mathbf{Z}/3\mathbf{Z}$	4	$[1,1,1,1,4,16]$	$\mathbf{Z}/4\mathbf{Z}$
9	$[1,1,1,1,10,10]$	(0), $\mathbf{Z}/5\mathbf{Z}$	11	$[1,1,1,2,3,16]$	(0), $\mathbf{Z}/2\mathbf{Z}$
13	$[1,1,1,2,5,14]$	(0), $\mathbf{Z}/2\mathbf{Z}$	27	$[1,1,1,5,6,10]$	(0), $\mathbf{Z}/2\mathbf{Z}$
31	$[1,1,2,2,2,16]$	$\mathbf{Z}/4\mathbf{Z}$	32	$[1,1,2,2,3,15]$	(0), $\mathbf{Z}/3\mathbf{Z}$
35	$[1,1,2,2,6,12]$	$\mathbf{Z}/2\mathbf{Z}$, $\mathbf{Z}/6\mathbf{Z}$	37	$[1,1,2,2,9,9]$	(0), $\mathbf{Z}/3\mathbf{Z}$
38	$[1,1,2,3,3,14]$	(0), $\mathbf{Z}/2\mathbf{Z}$	44	$[1,1,2,4,4,12]$	$\mathbf{Z}/4\mathbf{Z}$
48	$[1,1,2,4,8,8]$	$\mathbf{Z}/8\mathbf{Z}$	53	$[1,1,3,3,4,12]$	$\mathbf{Z}/3\mathbf{Z}$, $\mathbf{Z}/6\mathbf{Z}$
55	$[1,1,3,3,8,8]$	(0), $\mathbf{Z}/2\mathbf{Z}$	69	$[1,2,2,3,4,12]$	$\mathbf{Z}/2\mathbf{Z}$, $\mathbf{Z}/4\mathbf{Z}$
92	$[1,3,4,4,4,8]$	$\mathbf{Z}/4\mathbf{Z}$			

Moreover, all the above possibilities for $MW(f)$ *in each of these* 17 *types are realizable.*

Once we have settled the problem on $MW(f)$, we next consider Question 0.1. Our result is the following:

Theorem 0.3. *Let* $f : X \to \mathbf{P}^1$ *be an extremal semi-stable elliptic* $K3$ *surface. If* $\sharp(MW(f)) \geq 4$*, then Question* 0.1 *has a positive answer except for* $m = 49$*.*

Remark 0.4. Let ϕ be the homomorphism from $MW(f)$ to $\mathbf{Z}/n_1\mathbf{Z} \times \cdots \times \mathbf{Z}/n_6\mathbf{Z}$ given in [**MP3**, §2], i.e., $\phi(s) = (a_1, \dots, a_6)$, where a_i is the component number of the irreducible component that s hits at the corresponding singular fiber. Since ϕ is injective, we can identify $MW(f)$ with its image by ϕ. Then we have:

(1) Let $g_m : Y_m \to \mathbf{P}^1$ be any Jacobian elliptic fibration of type m with $MW(g_m) = (0)$ and fitting one of the nine cases in Theorem 0.2. Let $\{I_{n_1}, I_{n_2}, \dots, I_{n_k}, I_{n_{k+1}}, \dots, I_{n_6}\}$ be the set of types of singular fibers of g_m so that $1 = n_1 = n_2 = \cdots = n_{k-1} < n_k \leq n_{k+1} \leq \cdots \leq n_6$. Then the Picard lattice $\operatorname{Pic} Y_m$ is identical to $U \oplus A_{n_k-1} \oplus \cdots \oplus A_{n_6-1}$ with the $\mathbf{Q}/2\mathbf{Z}$-valued discriminant quadratic form $q_{\operatorname{Pic} Y_m}$ equal to (cf. [**Mo**]):

$$(-(n_k-1)/n_k) \oplus \cdots \oplus (-(n_6-1)/n_6).$$

Here $U = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, and the dual $(\operatorname{Pic} Y_m)^\vee = \operatorname{Hom}_{\mathbf{Z}}(\operatorname{Pic} Y_m, \mathbf{Z})$ naturally contains $\operatorname{Pic} Y_m$ as a sublattice with $\mathbf{Z}/n_k\mathbf{Z} \oplus \cdots \oplus \mathbf{Z}/n_6\mathbf{Z}$ as the factor group (see §1 for definitions).

An easy case-by-case check, using the fact that $q_{(T_{Y_m})} = -q_{(\mathrm{Pic}\, Y_m)}$, shows that the intersection matrix of the transcendental lattice T_{Y_m} is, modulo the action of $SL_2(\mathbf{Z})$, uniquely determined by the data $[n_1, \dots, n_6]$ (see [**Ni**, Prop. 1.6.1] or [**Mo**, Lemma 2.4]). So the intersection matrix of T_{Y_m} is equal to the corresponding one in the proof of Lemma (3.3). Thus, for each of these 9 of type m, there is exactly one $K3$ surface (modulo isomorphisms of abstract surfaces without the fibered structure being taken into consideration) which has a Jacobian elliptic fibration of type m with trivial Mordell-Weil group.

Also, for both $(m, G_m) = (35, \mathbf{Z}/2\mathbf{Z}), (53, \mathbf{Z}/3\mathbf{Z})$, there is a unique $K3$ surface X_m, which has a Jacobian elliptic fibration f_m of type m and $MW(f_m) = G_m$, because we can prove that the transcendental lattice T_{X_m} is unique in each pair case and identical to the corresponding one in the proof of Lemma (3.3).

The authors suspect that if $(f_m)_i : (X_m)_i \to \mathbf{P}^1$ are two Jacobian elliptic surfaces of the same type m and with $MW((f_m)_1) \cong MW((f_m)_2)$ then $(X_m)_1 \cong (X_m)_2$, though there may not be any fibered surface isomorphism between $((X_m)_i, (f_m)_i)$ $(i = 1, 2)$; see the fourth remark below and our Proposition (4.9). The importance of Lemma (3.3) is that its proof can be used, we guess, to lattice-theoretically show the existence of all cases of m and possibly to give an affirmative answer to this question. See [**SZ**] and [**Y**] for the non-semistable cases.

(2) When $m = 49$, we have $MW(f) = \mathbf{Z}/5\mathbf{Z}$ with $s_1 = (0,0,0,2,2,2)$ or $s_2 = (0,0,0,1,1,4)$ as its generator (cf. the Table in [**MP3**]). However, we have $2s_2 = (0,0,0,2,2,10-2)$. So we may assume that $MW(f)$ always has s_1 as its generator after suitable relabeling of fiber components if necessary.

(3) When $m = 110$, we have $MW(f) = \mathbf{Z}/3\mathbf{Z} \times \mathbf{Z}/3\mathbf{Z}$ with

$$G_1 = \{s_1 = (0,0,1,1,2,2), s_2 = (1,1,2,2,0,2)\}$$

or

$$G_2 = \{s_1 = (0,0,1,1,2,2), s_3 = (1,1,1,1,0,4)\}$$

as its set of generators (cf. the Table in [**MP3**]). Note that G_2 can be replaced by the new generating set $G_2' := \{s_1, 2s_3 = (3-1, 3-1, 2, 2, 0, 2)\}$. So we may assume that $MW(f)$ always has G_1 as its set of generators after suitable relabeling of fiber components if necessary.

(4) When $m = 46$, we have $MW(f) = \mathbf{Z}/2\mathbf{Z}$ with $s_1 = (0,0,0,0,3,5)$ or $s_2 = (0,0,1,2,0,5)$ as its generator (cf. the Table in [**MP3**]). As in the proof of Lemma (3.8), one can show that there are pairs (X_i, f_i) $(i = 1, 2)$ of the same type $m = 46$ with $MW(f_i) = \{O, s_i\}$. Moreover, the minimal resolution Y_i of $X_i/\langle s_i \rangle$ for $i = 1$ (resp. $i = 2$) has an

elliptic fibration $g_i : Y_i \to \mathbf{P}^1$, *induced from* f_i, of type $m = 101$ (resp. $m = 66$). Hence there is no isomorphism between the pairs (X_i, f_i).

(5) For $m = 69$, we have either $MW(f) = \mathbf{Z}/2\mathbf{Z}$ with $s = (0, 1, 1, 0, 0, 6)$ as its generator, or $MW(f) = \mathbf{Z}/4\mathbf{Z}$ with $s = (0, 1, 1, 0, 1, 3)$ as its generator (cf. Lemma (3.7)).

The contents of this article are as follows: In §1, we explain our technique and we give a brief summary of the facts we need. In §2, we give a method to construct (or show the nonexistence) of elliptic fibrations and give several examples of extremal elliptic $K3$ surfaces with trivial Mordell-Weil groups. §3 and §4 are devoted to proving Theorems 0.2 and 0.3, respectively.

Acknowledgment. Part of this work was done during the second author's visit to National University of Singapore (NUS) under the exchange program between NUS and the Japan Society of Promotion of Science (JSPS). Deep appreciation goes to both NUS and JSPS. The authors would like to thank Prof. S. Kondo for suggesting Lemma (3.1).

Conventions. In this article, the ground field is always the complex numbers field $\mathbf{C}$.

To describe the type of simple singularities of plane curves, we use bold capital letters, $\mathbf{A}$, $\mathbf{D}$ and $\mathbf{E}$.

We use capital italic letters A, D and E to describe the type of lattices, but we always multiply the value of intersection form by -1 for such lattices.

1. Preliminaries.

1.1. Cremona transformations and its applications.

We fix notation about Cremona transformations related with two-dimensional families of conics.

Let V be the vector space of homogeneous polynomials of degree 2 in three variables. Let $P, Q, R \in \mathbf{P}^2$ be three different points in general position and let $V_{P,Q,R}$ be the subspace of elements of V vanishing at P, Q and R; it is a 3-dimensional vector space. It is classical to define a rational map $CR_{P,Q,R}$: $\mathbf{P}^2 \dashrightarrow \check{\mathbf{P}}(V_{P,Q,R})$ where if $P_0 \in \mathbf{P}^2$, its image is the hyperplane of elements of $V_{P,Q,R}$ which also vanish at P_0. By a suitable choice of coordinates and the identification of $\check{\mathbf{P}}(V_{P,Q,R})$ with $\mathbf{P}^2$ this map may be written as:

$$\begin{array}{ccc} \mathbf{P}^2 & \dashrightarrow & \mathbf{P}^2 \\ [x : y : z] & \mapsto & [yz : xz : xy] \,. \end{array}$$

The map $CR_{P,Q,R}$ is not defined at P, Q, R, which are called the centers of the Cremona transformation. Outside the lines joining P, Q, R, this map is an isomorphism.

Let us consider now $P, Q \in \mathbf{P}^2$ and a line L through P such that $Q \notin L$. In the same way we define $V_{P,L,Q}$ as the space of equation of conics passing through P and Q and tangent to L at P. We define in the same way $CR_{P,L,Q}$. We can choose coordinates such that we have:

$$\begin{array}{ccc} \mathbf{P}^2 & \dashrightarrow & \mathbf{P}^2 \\ [x:y:z] & \mapsto & \left[y^2 : xy : xz\right]. \end{array}$$

This map is not defined at P and Q and it is an isomorphism outside L and the line joining P and Q. We say that the centers are Q and the two first infinitely near points of L at P; sometimes we will replace in the notation L by any curve through P whose only tangent at P is L.

There is a third type of Cremona transformation associated to a conic. Let C be a smooth conic passing through a point P; we denote $V_{P,C}$ as the space of equations of conics C' such that $(C \cdot C')_P = 3$. We denote $CR_{P,C}$ the associated Cremona transformation. It is not defined at P and is an isomorphism outside the tangent line to C at P. We say that the centers at P are the three first infinitely near points of C at P; sometimes we will replace in the notation Q by any curve through P such that Q is the only conic with highest contact at P. We can choose equations to write it down as:

$$\begin{array}{ccc} \mathbf{P}^2 & \dashrightarrow & \mathbf{P}^2 \\ [x:y:z] & \mapsto & \left[x^2 : xy : y^2 - xz\right]. \end{array}$$

1.2. Some lattice theory.

We here briefly review Nikulin's lattice theory. Details are found in [**Ni**]. Let L be a lattice, i.e.,

(i) L is a free finite $\mathbf{Z}$-module and
(ii) L is equipped with a nondegenerate bilinear symmetric pairing $\langle\ ,\ \rangle$.

For a given lattice L, $\operatorname{disc} L$ is the determinant of the intersection matrix. Note that it is independent of the choice of a basis. We call L unimodular if $\operatorname{disc} L = \pm 1$. Let J be a sublattice of L. We denote its orthogonal complement with respect to $\langle\ ,\ \rangle$ by $J^\perp$.

For a lattice L, we denote its dual lattice by $L^\vee$. Note that, by using the pairing, L is embedded in $L^\vee$ as a sublattice with same rank. Hence the quotient group $L^\vee/L$ is a finite Abelian group, which we denote by G_L.

L is called even if $\langle x, x\rangle$ is even for all $x \in L$. For an even lattice L, we define a quadratic form q_L with values in $\mathbf{Q}/2\mathbf{Z}$ as follows:

$$q_L(x \bmod L) = \langle x, x\rangle \bmod 2\mathbf{Z}.$$

Then we have the following lemma:

Lemma 1.1. *Let L be an even unimodular lattice. Let J_1 and J_2 be sublattices of L such that $J_1^\perp = J_2$ and $J_2^\perp = J_1$. Then*

(i) $G_{J_1} \cong G_{J_2}$ *and*
(ii) $q_{J_1} = -q_{J_2}$.

For a proof, see [**Ni**].

A sublattice M of L is called primitive if L/M is torsion-free.

Example 1.2. For a $K3$ surface X, $H^2(X, \mathbf{Z})$ is an even unimodular lattice with respect to the intersection pairing. The Picard group, $\operatorname{Pic} X$, is a primitive sublattice of $H^2(X, \mathbf{Z})$, and $T_X := (\operatorname{Pic} X)^\perp$ is called the transcendental lattice of X.

We shall end this subsection with the following lemma.

Lemma 1.3. *For $j = 1, 2$, let $\Delta_j = \Delta(1)_j \oplus \cdots \oplus \Delta(r_j)_j$ be a lattice where each $\Delta(i)_j$ is of Dynkin type A_a, D_d or E_e.*

(1) *Suppose that $\Phi : \Delta_1 \to \Delta_2$ is a lattice-isometry. Then $r_1 = r_2$ and $\Phi(\Delta(i)_1) = \Delta(i)_2$ after relabeling.*
(2) *Let $\mathbf{A} = A_{m_1} \oplus \cdots \oplus A_{m_k}$ be a direct sum of lattices of Dynkin type A_{m_i}. Suppose that $\mathbf{A}$ is an index-n $(n > 1)$ sublattice of $\Delta := \Delta_2$ and that $(m_1, \ldots, m_k) = (1, 1, 5, 11), (2, 2, 3, 11)$. Then one of the following three cases occurs (the first two are quite unlikely but the authors do not have a proof yet) :*
(2-1) $\mathbf{A} = A_1 \oplus (A_1 \oplus A_5 \oplus A_{11}), \Delta = A_1 \oplus D_{17}$, *and* $(A_1 \oplus A_5 \oplus A_{11}) \subseteq D_{17}$ *is an index-*6 *extension.*
(2-2) $\mathbf{A} = A_2 \oplus (A_2 \oplus A_3 \oplus A_{11}), \Delta = A_2 \oplus D_{16}$, *and* $(A_2 \oplus A_3 \oplus A_{11}) \subseteq D_{16}$ *is an index-*6 *extension.*
(2-3) $\mathbf{A} = A_1 \oplus A_{11} \oplus (A_1 \oplus A_5), \Delta = A_1 \oplus A_{11} \oplus E_6$, *and* $(A_1 \oplus A_5) \subseteq E_6$ *is an index-*2 *extension.*

Proof. We observe that

$$|\det(A_n)| = n + 1, \quad |\det(D_n)| = 4, \quad |\det(E_6)| = 3,$$
$$|\det(E_7)| = 2, \quad |\det(E_8)| = 1.$$

We also note that for an index n lattice extension $L \subseteq M$ one has $|\det(L)| = n^2 |\det(M)|$.

(1) is true when $r_1 = r_2 = 1$. In general, for a generating root e in $\Delta(1)_1$ with $e^2 = -2$, one has $(\Phi(e))^2 = -2$ and hence $\Phi(e) \in \Delta(1)_2$ say, because Δ_2 is even and negative definite. Now the connectedness of $\Delta(1)_1$ implies that $\Phi(\Delta(1)_1) \subseteq \Delta(1)_2$. Thus to prove (1), we may assume that $r_2 = 1, \Delta_2 = \Delta(1)_2$. The same argument applied to Φ^{-1} shows that $r_1 = 1$.

(2) The argument in (1) applied to the inclusion $\mathbf{A} \hookrightarrow \Delta_2$, implies that each $\Delta(i)_1$ contains a finite-index sublattice which is a sum of a few summands of $\mathbf{A}$. Now it follows from the observations at the beginning of the proof of this lemma, that either (2) is true or one of the following two cases occurs:

Case (2-4) $\mathbf{A} = A_{11} \oplus (A_2 \oplus A_2 \oplus A_3), \Delta = A_{11} \oplus D_7$, and $(A_2 \oplus A_2 \oplus A_3) \subseteq D_7$ is an index-3 extension.

Case (2-5) $\mathbf{A} = A_2 \oplus A_3 \oplus (A_2 \oplus A_{11}), \Delta = A_2 \oplus A_3 \oplus D_{13}$, and $(A_2 \oplus A_{11}) \subseteq D_{13}$ is an index-3 extension.

In the following, if e_i's form a canonical $\mathbf{Z}$-basis of A_n we let $h_n = (1/(n+1)) \sum_{i=1}^{n} ie_i \pmod{A_n}$ be the generator of $(A_n)^\vee / A_n \cong \mathbf{Z}/(n+1)\mathbf{Z}$. Note that $(h_n)^2 = -n/(n+1)$.

Suppose the contrary that Case (2-4) occurs. Set $\mathbf{B} = A_2 \oplus A_2 \oplus A_3$. Then $D_7 \subseteq \mathbf{B}^\vee := \mathrm{Hom}_{\mathbf{Z}}(\mathbf{B}, \mathbf{Z})$. and the latter is generated by h_2, h_2', h_3 with $(h_2)^2 = -2/3 = (h_2')^2, (h_3)^2 = -3/4$. Since D_7 is generated by roots and contains an index-3 sublattice $\mathbf{B}$, there is a root $t \in D_7 - \mathbf{B}$, and we can write $t = ah_2 + bh_2' + A$ where $a, b \in \mathbf{Z}, A \in \mathbf{B}$. Then $-2 = t^2 = (-2/3)(a^2 + b^2) + A^2 - 2s_1$ for some $s_1 \in \mathbf{Z}$. Since $\mathbf{B}$ is even and negative definite, $A^2 = -2s_2$ for some $s_2 \in \mathbf{Z}$. Denote by $s = s_1 + s_2$. Then $3 = a^2 + b^2 + 3s$, $3|(a^2 + b^2)$. Hence $a = 3a_1, b = 3b_1$ for some $a_1, b_1 \in \mathbf{Z}$. This leads to that $t = a_1(3h_2) + b_1(3h_2') + A \in \mathbf{B}$, a contradiction.

Suppose the contrary that Case (2-5) occurs. Set $\mathbf{B} = A_2 \oplus A_{11}$. Then $D_{13} \subseteq \mathbf{B}^\vee$ and the latter is generated by h_2, h_{11}. As in Case (2-4), there is a root $t \in D_{13} - \mathbf{B}$, and we can write $t = ah_2 + 4bh_{11} + A$ where $a, b \in \mathbf{Z}, A \in \mathbf{B}$. Then $-2 = t^2 = (-2/3)(a^2 + 22b^2) - 2s$ for some $s \in \mathbf{Z}$. Hence $3 = a^2 + 22b^2 + 3s$, $3|(a^2 + b^2)$ and $a = 3a_1, b = 3b_1$ for some $a_1, b_1 \in \mathbf{Z}$. This leads to that $t \in \mathbf{B}$, a contradiction. ☐

1.3. Review on elliptic surfaces with large torsion group.

We here give a brief summary on the results in [**CP**] and [**C**]. Let $f : X \to C$ be an elliptic surface over a curve C with a section O. Let $MW(f)$ be its Mordell-Weil group, the group of sections, O being the zero element. We denote its torsion part by $MW(f)_{\mathrm{tor}}$. Suppose that $MW(f)_{\mathrm{tor}} \supset \mathbf{Z}/m\mathbf{Z} \oplus \mathbf{Z}/n\mathbf{Z}$, $m|n$, $mn \geq 3$, and the j-invariant of X is not constant. Then it is known that one obtains $f : X \to C$ in a certain universal way, which we describe below. For that purpose, we need some notations.

Set

$$\Gamma_m(n) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbf{Z}) \mid \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \bmod n,\, b \equiv 0 \bmod m \right\}.$$

Let $X_m(n) = \Gamma_m(n) \backslash \mathcal{H}^*$, where $\mathcal{H}$ is the upper halfplane in $\mathbf{C}$, and let $E_m(n)$ be the elliptic modular surface of $\Gamma_m(n)$. By definition, $E_m(n)$ is an elliptic surface over $X_m(n)$; and we denote the morphism from $E_m(n)$ to $X_m(n)$ by $\psi_{m,n}$.

Suppose that $MW(f)_{\rm tor} \supset \mathbf{Z}/m\mathbf{Z} \oplus \mathbf{Z}/n\mathbf{Z}$, $m|n$, $mn \geq 3$. Then we have a commutative diagram

$$\begin{array}{ccc} C & \xrightarrow{g} & X_1(N) \\ & j \searrow & \downarrow j_{m,n} \\ & & \mathbf{P}^1 \end{array}$$

where j and $j_{m,n}$ are the j-invariants of f and $\psi_{m,n}$, respectively. Moreover, this diagram essentially gives $f : X \to C$, i.e., X is obtained as the pull-back surface by g, in the sense of relatively minimal smooth model.

Thus f is determined by g. Hence the uniqueness of X is reduced to that of g, which we consider in §4.

1.4. Comments on pencil of plane curves and nodal cubics.

Let $C = \{f = 0\}$ and $D = \{g = 0\}$ two projective plane curves of degree d without common components. They define a pencil of curves by considering $\{C_{[t:s]}\}_{[t:s]\in\mathbf{P}^1}$, where $C_{[t:s]}$ is the curve of equation $sf - tg = 0$. Let us denote $\mathcal{B} := C \cap D$; it is the set of base points of the pencils; these base points are the intersection points of any couple of elements of the pencil. A base point P is multiple if $(C \cdot D)_P > 1$ (we may replace C and D by any couple of different elements of the pencil). A pencil defines a rational map $\mathbf{P}^2 \dashrightarrow \mathbf{P}^1$ which is well-defined outside the base points. Let $Z \subset \mathbf{P}^2$ be an irreducible curve of degree e which is not a component of any element in the pencil. Let $C_{[t:s]}$ a generic element of the pencil. Then the pencil defines a map $\phi : Z \to \mathbf{P}^1$ of degree

$$d_Z := de - \sum_{P\in\mathcal{B}} (Z \cdot C_{[t:s]})_P;$$

if a base point P is in Z its image is the unique value $\phi(P)$ such that $(Z \cdot C_{\phi(P)})_P$ is greater than the generic intersection number. The critical points of the map are the points $Q \in Z$ such that:

- If Q is not a base point, then $C_{\phi(Q)}$ is either singular at Q or not transversal to Z at Q, i.e., $(Z \cdot C_{\phi(Q)})_Q > 1$.
- If $Q \in \mathcal{B}$, then $(Z \cdot C_{\phi(Q)})_Q > 1 + (Z \cdot C_{[t:s]})_P$, for $[t : s] \neq \phi(Q)$.

Let us consider a nodal cubic N in $\mathbf{P}^2$. We will apply later the following well-known result.

Proposition 1.4. *There exists a homogeneous coordinate system* $[x : y : z]$ *in* $\mathbf{P}^2$ *such that the equation of* N *is* $xyz + x^3 - y^3 = 0$. *The subgroup* G *of* $PGL(3, \mathbf{C})$ *fixing* N *is isomorphic to the dihedral group of order* 6. *Let* $\varphi : \mathbf{C}^* \to \mathrm{Reg}(N)$ *be the mapping defining by* $\varphi(t) := [t : t^2 : t^3 - 1]$. *Let us consider on* N *the geometrical group structure with zero element* $[1 : 1 : 0] = \varphi(1)$. *Then* φ *is a group isomorphism. Each element of* G *is determined by its action on* $\mathrm{Reg}(N)$*; the induced action on* $\mathbf{C}^*$ *is generated by* $t \mapsto t^{-1}$ *and* $t \mapsto \zeta t$ *where* $\zeta^3 = 1$.

2. Some extremal elliptic $K3$ surfaces with trivial Mordell-Weil group.

2.1. Elliptic fibrations and sextic curves.

Relationship between extremal elliptic fibrations and maximizing sextic curves was intensively studied in Persson's paper [**P**]. We explain in this section how to apply this method to construct or discard extremal elliptic fibrations. Let (X, f) be a pair such that X is a $K3$ surface and $f : X \to \mathbf{P}^1$ is a relatively minimal elliptic fibration with a fixed section O.

Step 1. Fix O as the zero element of the Mordell-Weil group $MW(f)$. It determines a group law on each regular fiber and it extends to a group law in the regular part of any fiber. For a fiber F of type I_n, there is a short exact sequence

$$0 \to \mathbf{C}^* \to \operatorname{Reg}(F) \to \mathbf{Z}/n\mathbf{Z} \to 0$$

where the kernel corresponds to the part of $\operatorname{Reg}(F)$ in the irreducible component which intersects O.

Step 2. On the regular part of any fiber F we can consider the map $P \mapsto -P$, (where $F \cap O$ is the zero element). These maps are the restriction of a morphism $\sigma : X \to X$, which is clearly an involution. By definition $f \circ \sigma = f$. Then, there is a natural map $\widetilde{\rho} : X/\sigma \to \mathbf{P}^1$; if F is an elliptic fiber of π, F/σ is the quotient of an elliptic curve by an involution with four fixed points (the 2-torsion), i.e., a smooth rational curve.

Then $\widetilde{\rho} : X/\sigma \to \mathbf{P}^1$ is a morphism from a smooth (rational) surface onto $\mathbf{P}^1$ whose generic fiber is $\mathbf{P}^1$. If F is a fiber of type I_{2n+1} (resp. I_{2n}), F/σ is a curve with normal crossings and $n+1$ irreducible components which are smooth and rational.

Step 3. For any singular fiber F, we contract all of the irreducible components of $\widetilde{\rho}(F)$ but the one which intersects $\widetilde{\rho}(O)$. We obtain a holomorphic fiber bundle $\rho : \Sigma \to \mathbf{P}^1$ with fiber isomorphic to $\mathbf{P}^1$ (Σ smooth) and a map $\tau : X \to \Sigma$ such that $\rho \circ \tau = \pi$. This map is generically $2:1$.

The map τ is a 2-fold covering ramified on the image of the fixed points of σ, i.e., on the image of the 2-torsion. We can write this curve as $E \cup R$ where $E := \tau(O)$, $R \cap E = \emptyset$ and R has intersection number three with the fibers of ρ. The number of irreducible components of R depends on the 2-torsion $T_2(MW(f))$ of the Mordell-Weil group of X (one irreducible component if $T_2(MW(f)) = 0$, two if $T_2(MW(f)) = \mathbf{Z}/2\mathbf{Z}$ and three if $T_2(MW(f)) = \mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/2\mathbf{Z}$).

If the configuration of π is $[1, \ldots, n_1, \ldots, n_r]$, $1 < n_1 \leq \cdots \leq n_r$, then R has exactly r singular points of type $\mathbf{A}_{n_1-1}, \ldots, \mathbf{A}_{n_r-1}$.

Remark 2.1. Let us suppose that $n_r > 7$, and let us call F the fiber of ρ containing this point $\mathbf{A}_{n_r-1}$; R intersects also F at another point P. Then we can perform three Nagata elementary transformations on the first three infinitely near points of R at $\mathbf{A}_{n_r-1}$. We call Σ' the result of this operation and we do not change the notation for the strict transforms; it induces a new fibration $\rho' : \Sigma' \to \mathbf{P}^1$ where E is a section of self-intersection -1. The curve R has a singular point $\mathbf{A}_{n_r-7}$ and $(R \cdot E)_P = 3$, and R is smooth at P. We can contract E and we obtain a projective plane where the contraction of R is a curve of degree 6 (also denoted by R) which has $r+1$ singular points of type $\mathbf{A}_{n_1-1}, \mathbf{A}_{n_2-1}, \ldots, \mathbf{A}_{n_r-7}$ and $\mathbf{E}_6$; the image of F is the tangent line to R at $\mathbf{E}_6$ and passes through $\mathbf{A}_{n_1-7}$. The pencil which induces the elliptic fibration (the *preferred pencil*) is the pencil of lines through $\mathbf{E}_6$. This fibration is called the standard fibration in [**P**] and in this case $\mathbf{E}_6$ is its center.

We can consider some kind of converse of this construction. Let $R \subset \mathbf{P}^2$ be a reduced curve (maybe reducible) of degree six such that its singular points are simple. Let P be a singular point of R. Then if X is the minimal resolution of the ramified double covering of $\mathbf{P}^2$ ramified on R and $\pi : X \to \mathbf{P}^1$ is the mapping induced by the pencil of lines through P, then π is a relatively minimal elliptic fibration of the $K3$-surface X. We call (X, π) the elliptic fibration associated to (R, P) and we will call the pencil of lines at P the preferred pencil; we will denote $\sigma : X \to \mathbf{P}^2$ the double covering. The following result is easy and useful.

Proposition 2.2. *Let $\pi : X \to \mathbf{P}^1$ be the elliptic fibration associated to (R, P) as above. Let E be a section of X; let $C := \sigma(E)$. Then either C is an irreducible component of R, either the intersection number of C and E at any point in $C \cap R$ is an even number.*

In both cases C is a curve of degree d having at P a singular point of multiplicity $d-1$. In the first case there is exactly one section over C and in the second case there are exactly two such sections.

We study now the existence of elliptic fibrations with trivial Mordell-Weil group in the cases of ambiguity which appear in the list of Miranda and Persson. In fact, we have applied this method to all cases of ambiguity in the list. As it is very long, we present only a few cases, where interesting phenomena occur.

2.2. Type $m = 9$.

Proposition 2.3. *There exist elliptic $K3$ surfaces of type 9, i.e., with configuration $[1, 1, 1, 1, 10, 10]$, and trivial Mordell-Weil group.*

This proposition gives one ambiguity case as such a fibration with Mordell-Weil group of order 5 appears in [**MP3**].

We look for an irreducible curve R of degree 6 having three singular points of type $\mathbf{E}_6, \mathbf{A}_3, \mathbf{A}_9$ and such that the tangent line to R at $\mathbf{E}_6$ passes through

$\mathbf{A}_3$. As in the case above the line through $\mathbf{A}_3$ and $\mathbf{A}_9$ intersects R at two other points.

Step 1. *First Cremona transformation.*

We consider $CR_{\mathbf{E}_6,\mathbf{A}_3,\mathbf{A}_9}$. We denote R_1 the strict transform of R; R_1 is a quintic curve. We have a smooth point Q such that the tangent line T to R_1 at Q verifies that $(R_1 \cdot Q)_Q = 4$. We denote Q' the other point in $R_1 \cap T$.

The other singular points of R_1 are $\mathbf{A}_7$ (coming from $\mathbf{A}_9$), P_1 (an ordinary double point coming from $\mathbf{A}_3$) and another ordinary double point denote P_2. The preferred pencil of lines has its center at P_1. The line joining P_1 and P_2 intersects R_1 at Q. The line joining P_1 and $\mathbf{A}_7$ passes through Q'. The ramification locus is $R_1 \cup T$.

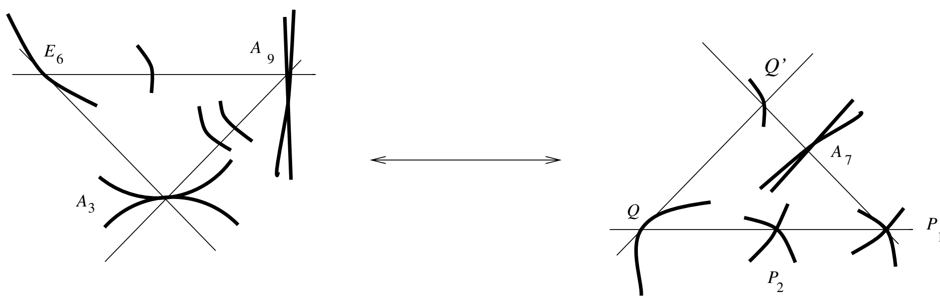

Figure 1.

Step 2. *Second and third Cremona transformations.*

We perform $CR_{P_1,P_2,\mathbf{A}_7}$. We obtain a quartic curve R_2 with one singular point $\mathbf{A}_5$ (coming from $\mathbf{A}_7$). The line T becomes a conic T_2 and $R_2 \cap T_2 = \{Q, Q', Q''\}$ where $(R_2 \cdot T_2)_Q = 5$, $(R_2 \cdot T_2)_{Q'} = 2$, $(R_2 \cdot T_2)_{Q''} = 1$, and $\mathbf{A}_5, Q', Q''$ are aligned. The center of the preferred pencil is Q''.

We perform the third Cremona transformation $CR_{\mathbf{A}_5,L,Q''}$, L being the tangent line at $\mathbf{A}_5$. We obtain two cubics R_3 and T_3. The cubic R_3 has an ordinary double point $\mathbf{A}_1$ and T_3 has also a double point denoted S (which is the center of the preferred pencil). The curves R_3 and T_3 have two intersection points Q and Q', with intersection numbers 5 and 4, and the points Q', S and $\mathbf{A}_1$ are aligned.

Question 2.4. Does there exist an irreducible nodal cubic R_3 (with node $\mathbf{A}_1$), an irreducible cubic T_3 with a double point S in $\mathbf{P}^2$ such that $R_3 \cap T_3 = \{Q, Q'\}$, $Q, Q' \neq S, \mathbf{A}_1$, with $(R_3 \cdot T_3)_Q = 5$, $(R_3 \cdot T_3)_{Q'} = 4$ and $Q', S, \mathbf{A}_1$ aligned?

Proposition 2.5. *The answer to Question* 2.4 *is yes.*

Proof. We proceed by applying Proposition 1.4 to R_3. We suppose that $Q = p(s^{-4})$ and $Q' = p(s^5)$. In this situation the equation of the line joining Q' and $\mathbf{A}_1$ is $y = s^5 x$. Let $f(x, y, z) = 0$ an equation for T_3 such that the coefficient of z^3 in f is 1. Then $f(t, t^2, t^3 - 1) = (t - s^5)^4 (t - s^{-4})^5$. We impose that T_3 intersects the line $y = s^2 x$ at one point outside Q' (with multiplicity 2). We force this point to be singular and we get the conditions on s (again with Maple-V). We obtain that

$$(s^6 - 1)(s^6 + 3s^3 + 1)(s^{12} + 4s^9 + s^6 + 4s^3 + 1) = 0.$$

We consider the action of the dihedral group; in the first term it is enough to retain the cases $s = \pm 1$; the positive case is too degenerate so it remains only $s = -1$. The equation of T_3 in this case is:

$$13\,y^3 + 9\,y^2 x - 5\,y^2 z - 9\,y x^2 - 6\,y x z - y z^2 - 13\,x^3 - 5\,x^2 z + x z^2 + z^3 = 0.$$

For the second term, one can see that we force $S = \mathbf{A}_1$ which is also too degenerate. The last factor gives two different cases (the twelve roots give two orbits by the action of the dihedral group). The equation is:

$$\begin{aligned}
&\left(-\frac{1265\,s^9}{2} - 60\,s^3 - \frac{4671}{2} - 2170\,s^6\right) x^3 \\
&+ \left(1205\,s^8 + 320\,s^{11} + 1285\,s^2\right) z x^2 \\
&+ \left(10080\,s + 135\,s^4 + 9480\,s^7 + 2466\,s^{10}\right) y x^2 \\
&+ \left(60\,s + 60\,s^7 + 16\,s^{10} + 5\,s^4\right) z^2 x \\
&+ \left(15255\,s^2 + 216\,s^5 + 14325\,s^8 + 3780\,s^{11}\right) y^2 x \\
&+ \left(\frac{495\,s^9}{2} + \frac{2103}{2} + 990\,s^6\right) y z x \\
&+ \left(-\frac{1735\,s^9}{2} - 60\,s^3 - \frac{6609}{2} - 3110\,s^6\right) y^3 \\
&- \left(640\,s + 620\,s^7 + 160\,s^{10} + 5\,s^4\right) z y^2 \\
&+ \left(-75\,s^2 - 75\,s^8 - 20\,s^{11} - 4\,s^5\right) z^2 y + z^3 = 0.
\end{aligned}$$

□

We deduce that there are essentially three different answers to Question 2.4. The main feature of the first answer is that the tangent line L to R_3 at Q' passes through Q. The elliptic surface is obtained from the double covering of $\mathbf{P}^2$ ramified along $R_3 + T_3$, and the elliptic fibration comes from the pencil of lines with center at S. One of the singular fibers is produced by the line joining S, $\mathbf{A}_1$ and Q'.

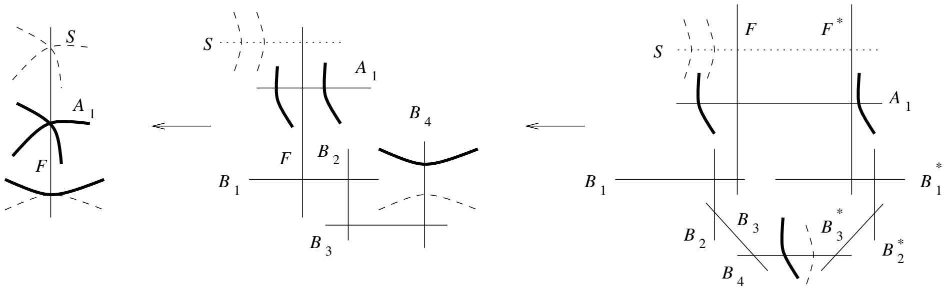

Figure 2.

The other singular fiber is produced by the line joining S and Q.

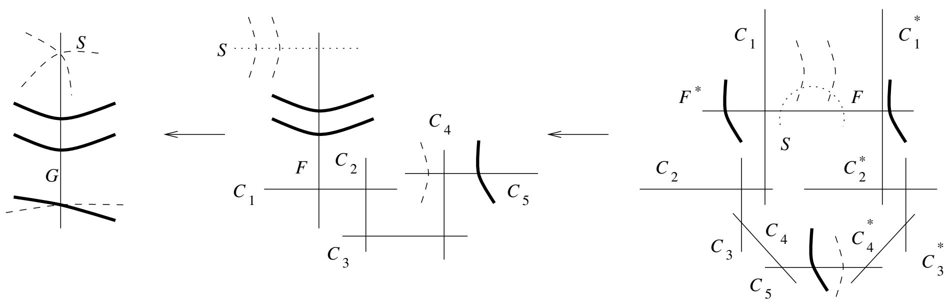

Figure 3.

Proposition 2.6. *The solution for $s = -1$ produces the elliptic fibration such that MW is cyclic of order 5. The solutions $s^{12}+4s^9+s^6+4s^3+1 = 0$ produce elliptic fibrations with trivial Mordell-Weil group; this case was not previously known.*

Proof. We note that the exceptional curve of the blowing-up of S never produces a section. In both cases the strict preimage of T_3 produces a section.

In the case $s = -1$, the intersection numbers of the line T with the curve R_3+T_3 are always even; then the preimage of L is reducible and produces two sections. We note also that Q is in this case an inflection point for both R_3 and T_3; the common tangent line has also even intersection numbers with $R_3 + T_3$ and then it produces two sections. We have found five different sections, then all of them.

Let us consider now the second case. We know already a section. By Proposition 2.2, any other section should come from a section to the pencil of lines through S having always even intersection numbers with the ramification curve $R_3 + T_3$. Then the problem is as follows:

Is there a curve D of degree d having a point of multiplicity $d-1$ at S and such that $(S \cdot R_3)_P \equiv (S \cdot T_3)_P \mod 2$ for any $P \in \mathbf{P}^2$ and any branch of D at S has even intersection number with T_3?

Let us suppose that such a curve exists. It gives two different sections D_0 and D_1 in the elliptic surface. From [**MP3**], D_0 and D_1 are torsion sections, and then they must be disjoint. In particular, D cannot intersect $R_3 \cup T_3$ outside $S, \mathbf{A}_1, Q, Q'$ and no branch of D at S is tangent to any branch of T_3 at S.

D_0 and D_1 belong to the 5-torsion, so by the structure of the singular fibers, we have:

– $\mathbf{A}_1 \notin D$;
– $(T_3 \cdot D)_{Q'} = (R_3 \cdot D)_{Q'} = a = 0, 2, 4$;
– $(T_3 \cdot D)_{Q'} = (R_3 \cdot D)_{Q} = b = 1, 3, 5$.

Then, putting all these conditions together, we obtain that $S \notin D$ and so D is a line; then $3 = a + b$. The two possibilities appear in the previous case, but not in this one. □

2.3. Case $m = 11$.

The method to find or discard the fibrations in the other cases is the same one. As the answers are positive, we will give the results that may be verified by the reader. Let us consider the polynomial

$$\begin{aligned} &p_1(x,y,z)\\ &:= \left(\frac{11593}{95004009} - \frac{4027\,v}{190008018}\right) y^4x^2 + \left(\frac{4705}{10556001} - \frac{2183\,v}{10556001}\right) zxy^4 \\ &+ \left(-\frac{1493\,v}{4691556} + \frac{803}{2345778}\right) z^2y^4 + \left(-\frac{48226}{5000211} + \frac{1475\,v}{5000211}\right) zy^3x^2 \\ &+ \left(\frac{1174\,v}{185193} - \frac{4736}{185193}\right) z^2xy^3 + \left(\frac{635\,v}{123462} - \frac{755}{61731}\right) z^3y^3 \\ &+ \left(\frac{20153}{87723} + \frac{1081\,v}{175446}\right) z^2y^2x^2 + \left(\frac{854}{3249} - \frac{187\,v}{3249}\right) z^3y^2x \\ &+ \left(-\frac{427}{6498} + \frac{187\,v}{12996}\right) z^4y^2 + \left(-\frac{22612}{13851} + \frac{386\,v}{13851}\right) z^3yx^2 \\ &+ \left(\frac{1412}{1539} + \frac{20\,v}{1539}\right) z^4xy + x^3z^3 + \left(-\frac{11\,v}{729} - \frac{485}{729}\right) z^4x^2 \end{aligned}$$

where $v^2 + 2 = 0$.

Proposition 2.7. *The curve $p_1(x, y, z) = 0$ is an irreducible curve with singularities $\mathbf{E}_6$ (at $[1:0:0]$ and tangent line $z = 0$), $\mathbf{A}_1$ (at $[0:0:1]$), $\mathbf{A}_9$ (at $[0:1:0]$) and $\mathbf{A}_2$ (at $[1:1:1]$). The pencil of lines through the* *triple point determine after a double covering an elliptic $K3$ fibration of type* $[1,1,1,2,3,16]$ *with trivial Mordell-Weil group.*

Proof. The computations have been performed with Maple-V. We note that the curve is irreducible as the line $x=0$ joining $\mathbf{A}_9$ and $\mathbf{A}_1$ is not a component. The Miranda-Persson classification finishes the result. □

2.4. Case $m = 13$.

Proposition 2.8. *The curve $p_2(x,y,z) = 0$ (see below) is an irreducible curve with singularities $\mathbf{E}_6$ (at $[1:0:0]$ and tangent line $y=0$), $\mathbf{A}_7$ (at $[0:0:1]$), $\mathbf{A}_4$ (at $[0:1:0]$) and $\mathbf{A}_1$ (at $[1:1:1]$). The pencil of lines through the triple point determine after a double covering an elliptic $K3$ fibration of type $[1,1,1,2,5,14]$ with trivial Mordell-Weil group.*

Proof. As before, computations have been performed with Maple-V. We note that the curve is irreducible as the line $x=y$ joining $\mathbf{A}_7$ and $\mathbf{A}_1$ is not a component. The Miranda-Persson classification finishes the result. □

We have:

$$
\begin{aligned}
&p_2(x,y,z)\\
&:= y^3x^3 + \left(-\frac{24284}{130321} + \frac{10287\,v}{260642} + \frac{144295\,v^2}{1824494}\right) y^4x^2\\
&+ \left(-\frac{6071515\,v^2}{130321} - \frac{2851308\,v}{130321} + \frac{13668817}{130321}\right) zx^2y^3\\
&+ \left(\frac{38660279\,v}{260642} + \frac{161684215\,v^2}{521284} - \frac{179634441}{260642}\right) z^2x^2y^2\\
&+ \left(-\frac{252208635\,v^2}{521284} - \frac{60782001\,v}{260642} + \frac{277127879}{260642}\right) z^3x^2y\\
&+ \left(\frac{55758423\,v}{521284} + \frac{460287135\,v^2}{2085136} - \frac{125694751}{260642}\right) z^4x^2\\
&+ \left(-\frac{10473}{6859} + \frac{2326\,v}{6859} + \frac{32860\,v^2}{48013}\right) zxy^4\\
&+ \left(-\frac{361050\,v^2}{6859} - \frac{176895\,v}{6859} + \frac{1579285}{13718}\right) z^2xy^3\\
&+ \left(\frac{725753\,v}{13718} + \frac{1458065\,v^2}{13718} - \frac{1564472}{6859}\right) z^3xy^2\\
&+ \left(\frac{1625477}{13718} - \frac{191737\,v}{6859} - \frac{3045105\,v^2}{54872}\right) z^4xy\\
&+ \left(-\frac{268}{361} + \frac{141\,v}{722} + \frac{3495\,v^2}{10108}\right) z^2y^4 + \left(\frac{825}{722} - \frac{255\,v}{361} - \frac{1175\,v^2}{1444}\right) z^3y^3
\end{aligned}
$$

$$+\left(-\frac{686}{361}+\frac{1099\,v}{1444}+\frac{6055\,v^2}{5776}\right)z^4y^2,$$

where $5\,v^3-4\,v^2-14\,v+14=0$.

Let us remark that this condition has exactly one real solution.

2.5. Case $m = 27$.

In this cases we only state the result concerning the existence and unicity of curves and we give the equation of the polynomial. The proofs and methods of computations are very similar to the previous ones.

Proposition 2.9. *The curve* $p_3(x, y, z) = 0$ *(see below) is an irreducible curve with singularities* $\mathbf{E}_6$ *(at* $[0:0:1]$ *and tangent line* $y = 0$*),* $\mathbf{A}_3$ *(at* $[1:0:0]$*),* $\mathbf{A}_5$ *(at* $[0:1:0]$*) and* $\mathbf{A}_4$ *(at* $[1:1:1]$*). The pencil of lines through the triple point determine after a double covering an elliptic* $K3$ *fibration of type* $[1, 1, 1, 5, 6, 10]$ *with trivial Mordell-Weil group.*

We have

$$\begin{aligned}
&p_3(x,y,z)\\
&:=\left(-\frac{200\,v^2}{297}-\frac{425}{297}-\frac{110\,v}{27}\right)y^4x^2+\left(\frac{125}{396}+\frac{5\,v}{9}-\frac{13\,v^2}{396}\right)zy^4x\\
&+\left(\frac{5\,z^2}{528}-\frac{5}{264}+\frac{5\,v}{48}\right)z^2y^4+\left(\frac{115\,v^2}{81}+\frac{220}{81}+\frac{875\,v}{81}\right)y^3x^3\\
&+\left(\frac{655}{108}+\frac{493\,v}{54}+\frac{133\,v^2}{108}\right)zy^3x^2+\left(\frac{5\,v^2}{36}-\frac{115}{36}-\frac{5\,v}{9}\right)z^2y^3x+z^3y^3\\
&+\left(-\frac{2225}{972}-\frac{3275\,v}{486}-\frac{725\,v^2}{972}\right)y^2x^4+\left(-\frac{2831}{324}-\frac{2032\,v}{81}-\frac{797\,v^2}{324}\right)zy^2x^3\\
&+\left(-\frac{37\,v^2}{72}-\frac{35}{36}-\frac{215\,v}{72}\right)z^2y^2x^2+\left(\frac{1225\,z^2}{972}+\frac{5215}{972}+\frac{7495\,v}{486}\right)zyx^4\\
&+\left(\frac{1105}{324}+\frac{788\,v}{81}+\frac{193\,v^2}{324}\right)z^2yx^3+\left(-\frac{893\,v^2}{3888}-\frac{4333}{1944}-\frac{24499\,v}{3888}\right)z^2x^4
\end{aligned}$$

where $25+75\,v+15\,v^2+v^3=0$.

2.6. Case $m = 32$.

Let us consider the polynomial

$$\begin{aligned}
&p_4(x,y,z)\\
&:=y^3z^3+\left(\frac{5625\,v}{668168}-\frac{33625}{334084}\right)z^2x^4+\left(\frac{3475\,v}{58956}+\frac{39275}{29478}\right)yz^2x^3\\
&+\left(-\frac{1465\,v}{1734}-\frac{1775}{867}\right)y^2x^2z^2+\left(\frac{173\,v}{204}-\frac{299}{102}\right)y^3xz^2
\end{aligned}$$

$$+\left(-\frac{v}{40}+\frac{17}{20}\right)y^4z^2+\left(\frac{19675\,v}{501126}-\frac{188825}{501126}\right)yzx^4$$
$$+\left(\frac{350\,v}{4913}+\frac{23110}{4913}\right)y^2x^3z+\left(-\frac{1580\,v}{867}-\frac{5900}{867}\right)y^3x^2z$$
$$+\left(\frac{11\,v}{15}-5/3\right)y^4xz+\left(\frac{29555\,v}{668168}-\frac{232705}{668168}\right)y^2x^4$$
$$+\left(-\frac{1885\,v}{29478}+\frac{116975}{29478}\right)y^3x^3+\left(-\frac{1205\,v}{1734}-\frac{33517}{8670}\right)y^4x^2$$

where $v^2-v+34=0$.

Proposition 2.10. *The curve $p_4(x,y,z)=0$ is an irreducible curve with singularities $\mathbf{E}_6$ (at $[0:0:1]$ and tangent line $y=0$), $\mathbf{A}_8$ (at $[1:0:0]$), $\mathbf{A}_2$ (at $[0:1:0]$) and two points of type $\mathbf{A}_1$ in the line $x+y+z=0$. The pencil of lines through the triple point determine after a double covering an elliptic $K3$ fibration of type $[1,1,2,2,3,15]$ with trivial Mordell-Weil group.*

2.7. Case $m=37$.

Proposition 2.11. *The curve $p_5(x,y,z)=0$ (see below) is an irreducible curve with singularities $\mathbf{E}_6$ (at $[0:0:1]$ and tangent line $x=0$), $\mathbf{A}_2$ (at $[0:1:0]$), $\mathbf{A}_8$ (at $[1:0:0]$) and two points of type $\mathbf{A}_1$ in the line $x+y+z=0$. The pencil of lines through the triple point determine after a double covering an elliptic $K3$ fibration of type $[1,1,2,2,9,9]$ with trivial Mordell-Weil group.*

We have:

$$p_5(x,y,z)$$
$$:=\left(\frac{3970803\,v}{130438}-\frac{345557847\,v^2}{65219}+\frac{8058927}{130438}\right)y^4x^2$$
$$+\left(-\frac{82574784\,v^2}{5929}+\frac{37159110\,v}{5929}-\frac{3105297}{5929}\right)zy^4x$$
$$+\left(-\frac{653967}{2156}+\frac{3545235\,v}{1078}-\frac{5380479\,v^2}{1078}\right)z^2y^4$$
$$+\left(\frac{5894214\,v}{9317}-\frac{295704\,v^2}{9317}-\frac{650011}{9317}\right)y^3x^3$$
$$+\left(-\frac{278076\,v^2}{847}+\frac{808926\,v}{847}-\frac{86286}{847}\right)zy^3x^2$$
$$+\left(-\frac{105723\,v^2}{77}+\frac{80505\,v}{77}-\frac{15255}{154}\right)z^2xy^3$$

$$+\left(\frac{14286}{1331}-\frac{136113\,v}{1331}+\frac{65742\,v^2}{1331}\right)y^2x^4$$
$$+\left(-\frac{24048\,v}{121}+\frac{30018\,v^2}{121}+\frac{4599}{242}\right)zy^2x^3$$
$$+\left(-\frac{2199\,v}{11}+\frac{3966\,v^2}{11}+\frac{195}{11}\right)z^2y^2x^2$$
$$+\left(-\frac{309}{121}+\frac{3711\,v}{121}-\frac{8358\,v^2}{121}\right)zyx^4$$
$$+\left(\frac{471\,v}{11}-\frac{903\,v^2}{11}-\frac{87}{22}\right)z^2yx^3+\left(-\frac{42\,v^2}{11}+\frac{159\,v}{44}-\frac{15}{44}\right)z^2x^4+z^3x^3$$

where $28\,v^3-30\,v^2+12\,v-1=0$.

2.8. Case $m = 38$.

Let us consider the polynomial

$$p_6(x,y,z)$$
$$:=\frac{1404\,x^2y^4}{1445}-\frac{9\,xy^4z}{85}+\frac{17\,z^2y^4}{60}+\frac{10800\,x^3y^3}{4913}+\frac{1980\,x^2y^3z}{289}$$
$$-\frac{37\,z^2y^3x}{102}+y^3z^3+\frac{105840\,x^4y^2}{83521}+\frac{4410\,x^3y^2z}{289}+\frac{13965\,z^2y^2x^2}{1156}$$
$$+\frac{720300\,x^4yz}{83521}+\frac{780325\,z^2yx^3}{29478}+\frac{14706125\,z^2x^4}{1002252}.$$

Proposition 2.12. *The curve $p_6(x,y,z)=0$ is an irreducible curve with singularities $\mathbf{E}_6$ (at $[0:0:1]$ and tangent line $y=0$), $\mathbf{A}_7$ (at $[1:0:0]$), $\mathbf{A}_1$ (at $[0:1:0]$) and two points of type $\mathbf{A}_2$ in the line $x+y+z=0$. The pencil of lines through the triple point determine after a double covering an elliptic $K3$ fibration of type $[1,1,2,3,3,14]$ with trivial Mordell-Weil group.*

2.9. Case $m = 55$.

Let us consider the polynomial

$$p_7(x,y,z)$$
$$:=\left(\frac{139}{176}+\frac{175\,v}{176}\right)y^4z^2+.\left(-\frac{837\,v}{242}+\frac{7101}{968}\right)y^4zx$$
$$+\left(\frac{30537}{10648}-\frac{29565\,v}{10648}\right)y^4x^2+\left(-\frac{151\,v}{44}+\frac{155}{44}\right)y^3z^2x$$
$$+\left(\frac{675}{242}+\frac{837\,v}{242}\right)y^3zx^2+\left(-\frac{669\,v}{2662}+\frac{2765}{1331}\right)y^3x^3$$
$$+\left(-\frac{81\,v}{22}+\frac{243}{44}\right)y^2z^2x^2+\left(\frac{441\,v}{242}-\frac{183}{242}\right)y^2zx^3$$

$$+\left(-\frac{1107}{1331}+\frac{2025\,v}{1331}\right)y^2x^4+\left(-\frac{17}{11}+\frac{107\,v}{22}\right)yz^2x^3$$
$$+\left(\frac{153\,v}{121}+\frac{18}{121}\right)yzx^4+z^3x^3+\left(\frac{13}{22}-\frac{5\,v}{22}\right)z^2x^4$$

where $3v^2-4v+2=0$.

Proposition 2.13. *The curve $p_7(x,y,z)=0$ is an irreducible curve with singularities $\mathbf{E}_6$ (at $[0:0:1]$ and tangent line $x=0$), $\mathbf{A}_1$ (at $[0:1:0]$), $\mathbf{A}_7$ (at $[1:0:0]$) and two points of type $\mathbf{A}_2$ in the line $x+y+z=0$. The pencil of lines through the triple point determine after a double covering an elliptic $K3$ fibration of type $[1,1,3,3,8,8]$ with trivial Mordell-Weil group.*

3. The complete determination of the Mordell-Weil group for each type of semi-stable extremal fibrations.

In this section, we shall show Theorem 0.2 which will follow from the Table in [**MP3**], and the Lemmas below. We recall Lemma 1.3 and Shioda-Inose's result that the isomorphism class of a $K3$ surface X of Picard number 20 is uniquely determined by the transcendental lattice T_X, modulo the action of $SL_2(\mathbf{Z})$ [**SI**].

Lemma 3.1. *Let S be an even symmetric lattice of rank 20 and signature $(1,19)$ and T a positive definite even symmetric lattice of rank 2. Assume that $\varphi: T^\vee/T \to S^\vee/S$ is an isomorphism which induces the following equality involving $\mathbf{Q}/2\mathbf{Z}$-valued discriminant (quadratic) forms: $q_S=-q_T$.*

Let X be the unique $K3$ surface (up to isomorphisms) with the transcendental lattice $T_X=T$. Then the Picard lattice $\operatorname{Pic} X$ is isometric to S.

Proof. Consider the overlattice L of $S\oplus T$ obtained by adding all elements $\varphi(x)+x$, $x\in T^\vee$, where $\varphi(x)\in S^\vee$ denotes one representative of $\varphi(x+T)\in S^\vee/S$. The (even) intersection form on $S\oplus T$ is naturally extended to a $\mathbf{Q}$-valued one on $S^\vee\oplus T^\vee$. For each $x\in T^\vee$, we have, modulo $2\mathbf{Z}$, $(\varphi(x)+x,\varphi(x)+x)=-q_T(x)+q_T(x)=0$, i.e., $(\varphi(x)+x,\varphi(x)+x)\in 2\mathbf{Z}$. Also for $x_i\in T^\vee$, combining $(\varphi(x_1+x_2),\varphi(x_1+x_2))=-(x_1+x_2,x_1+x_2)$ (mod $2\mathbf{Z}$) and $(\varphi(x_i),\varphi(x_i))=-(x_i,x_i)$ (mod $2\mathbf{Z}$), we see that $(\varphi(x_1),\varphi(x_2))=-(x_1,x_2)$ (mod $\mathbf{Z}$), whence mod $\mathbf{Z}$ we have $(\varphi(x_1)+x_1,\varphi(x_2)+x_2)=0$. Thus L is an even (integral) symmetric lattice of rank 22 and signature $(1+2,19+0)$. Clearly, $L/(S\oplus T)\cong T^\vee/T$ and hence $|\det(L)|=|\det(S\oplus T)|/|T^\vee/T|^2=1$. Now by the classification of indefinite unimodular even symmetric lattices, L is isometric to the $K3$ lattice (cf. [**Se**]).

On the other hand, by [**SI**], there is a unique $K3$ surface X (modulo isomorphisms) with the intersection form of the transcendental lattice T_X equal to T (modulo $SL_2(\mathbf{Z})$). We identify L with $H^2(X,\mathbf{Z})$ and T with T_X. Note that there are two embeddings $\iota_k: T_X\to H^2(X,\mathbf{Z})$: $\iota_1: T_X\hookrightarrow$

$H^2(X, \mathbf{Z})$ as the transcendental sublattice, and $\iota_2 : T_X = T \hookrightarrow S \oplus T \hookrightarrow L = H^2(X, \mathbf{Z})$.

The embedding ι_1 (resp. ι_2) is primitive by the definition of T_X (resp. of L). Now Nikulin's uniqueness theorem of primitive embedding implies that there is an isometry Ψ of $H^2(X, \mathbf{Z})$ such that $\iota_1 = \Psi \circ \iota_2$ [**Mo**, Cor. 2.10]. Note that the Picard lattice $\mathrm{Pic}X = (\iota_1(T_X))^{\perp} = (\Psi(\iota_2(T_X)))^{\perp} = \Psi(T^{\perp}) = \Psi(S) \cong S$. □

Lemma 3.2. *Let $f : X \to \mathbf{P}^1$ be of type $m = 4$ as in Theorem* 0.2. *Then $MW(f) \neq (0)$.*

Proof. Suppose the contrary that $f : X \to \mathbf{P}^1$ is of type $m = 4$ with $MW(f) = (0)$. Then Pic X is a direct sum $U \oplus A_3 \oplus A_{15}$ of lattices, where $U = (a_{ij})$ satisfies $a_{ii} = 0, a_{12} = a_{21} = 1$. Let (b_{ij}) be the intersection matrix of the transcendental lattice $T = T_X$. Then $b_{ii} > 0$ and $\det(b_{ij}) = |\det(\mathrm{Pic}\, X)| = 64$ (cf. [**BPV**]). After conjugation by $SL(2, \mathbf{Z})$, we may assume that $-b_{11} < 2|b_{12}| \le b_{11} \le b_{22}$, and that $b_{12} \ge 0$ when $b_{11} = b_{22}$. An easy calculation shows that one of the following cases occurs:

(1) $(b_{ij}) = \mathrm{diag}\,[2, 32]$,
(2) $(b_{ij}) = \mathrm{diag}\,[4, 16]$,
(3) $(b_{ij}) = \mathrm{diag}\,[8, 8]$, and
(4) $b_{11} = 8, b_{22} = 10, b_{12} = 4$.

Embed T, as a sublattice, naturally into $T^{\vee} = \mathrm{Hom}_{\mathbf{Z}}(T, \mathbf{Z})$. Then $T^{\vee}/T \cong (\mathrm{Pic}X)^{\vee}/\ (\mathrm{Pic}X) \cong \mathbf{Z}/4\mathbf{Z} \oplus \mathbf{Z}/16\mathbf{Z}$. Note that $(\mathrm{Pic}X)^{\vee}/(\mathrm{Pic}X)$ is generated by $\varepsilon_1 = (1/4)\sum_{i=1}^{3} i v_i$ and $\varepsilon_2 = (1/16)\sum_{i=4}^{18}(i-3)v_i$, modulo Pic X, where v_i's form a canonical basis of $A_3 \oplus A_{15} \subseteq$ Pic X. So the discriminantal quadratic form $q_T : T^{\vee}/T \to \mathbf{Q}/2\mathbf{Z}$ is equal to $-q_{\mathrm{Pic}\, X} = (-\varepsilon_1^2) \oplus (-\varepsilon_2^2) = (3/4) \oplus (15/16)$.

On the other hand, in Case (4), $T^{\vee}$ has a $\mathbf{Z}$-basis $(e_1\ e_2)(b_{ij})^{-1} = (g_1\ g_2)$, where e_1, e_2 form a canonical basis of T, where $g_1 = (1/32)(5e_1 - 2e_2), g_2 = (1/16)(-e_1 + 2e_2)$. This leads to that $\mathrm{ord}(g_1)$ is equal to 32 in $T^{\vee}/T$, a contradiction.

In Cases (1)-(3) where $T = \mathrm{diag}\,[s, t]$, with $(s, t) = (2, 32), (4, 16)$ or $(8, 8)$, the discriminantal quadratic form q_T is equal to $(1/s) \oplus (1/t)$. This leads to that $(1/s) \oplus (1/t) \cong (3/4) \oplus (15/16)$, which is impossible by an easy check. □

Lemma 3.3. *Consider the pairs below:*

$$(m, G_m) = (2, \langle 0 \rangle), (9, \langle 0 \rangle), (11, \langle 0 \rangle), (13, \langle 0 \rangle), (27, \langle 0 \rangle), (32, \langle 0 \rangle),$$
$$(37, \langle 0 \rangle), (38, \langle 0 \rangle), (55, \langle 0 \rangle), (35, \mathbf{Z}/2\mathbf{Z}), (53, \langle \mathbf{Z}/3\mathbf{Z} \rangle).$$

For each of these eleven pairs (m, G_m), there is a Jacobian elliptic $K3$ surface $f_m : X_m \to \mathbf{P}^1$ of type m as in Theorem 0.2 *such that $(m, MW(f_m)) = (m, G_m)$.*

Proof. The existence of the pairs where $m = 2, 35$ is proved constructively in [**AT**]. The rest is also constructively proved in §2. In the paragraphs below, we will give an independent lattice-theoretical proof.

Let T_m, $m = 2, 9, 11, 13, 27, 32, 37, 38, 55, 35, 53$, be the positive definite even symmetric lattice of rank 2 with the following intersection form, respectively:

$$\begin{pmatrix} 4 & 2 \\ 2 & 10 \end{pmatrix}, \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}, \begin{pmatrix} 10 & 2 \\ 2 & 10 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 70 \end{pmatrix}, \begin{pmatrix} 10 & 0 \\ 0 & 30 \end{pmatrix}, \begin{pmatrix} 12 & 6 \\ 6 & 18 \end{pmatrix},$$

$$\begin{pmatrix} 18 & 0 \\ 0 & 18 \end{pmatrix}, \begin{pmatrix} 6 & 0 \\ 0 & 42 \end{pmatrix}, \begin{pmatrix} 24 & 0 \\ 0 & 24 \end{pmatrix}, \begin{pmatrix} 6 & 0 \\ 0 & 12 \end{pmatrix}, \begin{pmatrix} 4 & 0 \\ 0 & 12 \end{pmatrix}.$$

For the first nine m above, let S_m be the even lattice of rank 20 and signature (1,19) with the following intersection form, respectively

$$U \oplus A_1 \oplus A_{17}, U \oplus A_9 \oplus A_9, U \oplus A_1 \oplus A_2 \oplus A_{15},$$

$$U \oplus A_1 \oplus A_4 \oplus A_{13}, U \oplus A_4 \oplus A_5 \oplus A_9, U \oplus A_1 \oplus A_1 \oplus A_2 \oplus A_{14},$$

$$U \oplus A_1 \oplus A_1 \oplus A_8 \oplus A_8, U \oplus A_1 \oplus A_2 \oplus A_2 \oplus A_{13}, U \oplus A_2 \oplus A_2 \oplus A_7 \oplus A_7.$$

We now define S_m for $m = 35, 53$. Let Γ_{35} be the lattice $U \oplus A_1 \oplus A_1 \oplus A_5 \oplus A_{11}$, with $G, H, J_i (1 \le i \le 5), \theta_i (1 \le i \le 11)$ as the canonical basis of $A_1 \oplus A_1 \oplus A_5 \oplus A_{11}$, and $\mathcal{O}, F$ as a basis of U such that $\mathcal{O}^2 = -2, F^2 = 0, \mathcal{O} \cdot F = 1$.

We extend Γ_{35} to an index-2 integral over lattice $S_{35} = \Gamma_{35} + \mathbf{Z} s_{35}$, where

$$s_{35} = \mathcal{O} + 2F - G/2 - H/2 - (1/2) \left(\sum_{i=1}^{6} i\theta_i + \sum_{i=7}^{11} (12 - i)\theta_i \right).$$

It is easy to see that the intersection form on Γ_{35} can be extended to an integral even symmetric lattice of signature $(1, 19)$. Indeed, setting $s = s_{35}$, we have

$$s^2 = -2, s \cdot F = s \cdot G = s \cdot H = s \cdot \theta_6 = 1, s \cdot \mathcal{O} = s \cdot J_i = s \cdot \theta_j = 0 \ (\forall i; j \ne 6).$$

Moreover, $|\det(S_{35})| = |\det(\Gamma_{35})|/2^2 = 72$.

Note that $\Gamma_{35}^\vee = \mathrm{Hom}_{\mathbf{Z}}(\Gamma_{35}, \mathbf{Z})$ contains naturally Γ_{35} as a sublattice with $\mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/6\mathbf{Z} \oplus \mathbf{Z}/12\mathbf{Z}$ as the factor group, and is generated by the following, modulo Γ_{35}:

$$h_1 = G/2, \ h_2 = H/2, \ h_3 = (1/6) \sum_{i=1}^{5} iJ_i, \ h_4 = (1/12) \sum_{i=1}^{11} i\theta_i.$$

Since $(S_{35})^\vee$ is an (index-2) sublattice of $(\Gamma_{35})^\vee$, an element x is in $(S_{35})^\vee$ if and only if $x = \sum_{i=1}^{4} a_i h_i$ (mod Γ_{35}) such that x is integral on S_{35}, i.e.,

$x \cdot s = (a_1 + a_2 + a_4)/2$ is an integer. Hence $(S_{35})^\vee$ is generated by the following, modulo Γ_{35}:

$$h_3,\ 2h_i,\ h_1 + h_2,\ h_1 + h_4,\ h_2 + h_4.$$

Noting that $2h_1, 2h_2 \in S_{35}$ and $(h_1 + h_2) + 6h_4$ is equal to s (mod Γ_{35}) and hence contained in S_{35}, we can see easily that $(S_{35})^\vee$ is generated by the following, modulo S_{35}:

$$\varepsilon_1 := h_3, \quad \varepsilon_2 := h_1 - h_4.$$

Now the fact that $|(S_{35})^\vee/S_{35}| = 72$ and that $6\varepsilon_1, 12\varepsilon_2 \in S_{35}$ imply that $(S_{35})^\vee/S_{35}$ is a direct sum of its cyclic subgroups which are of order 6, 12, and generated by $\varepsilon_1, \varepsilon_2$, modulo S_{35}.

We note that the negative of the discriminant form

$$-q_{(S_{35})} = (-(\varepsilon_1)^2) \oplus (-(\varepsilon_2)^2) = (5/6) \oplus ((1/2) + (11/12)) = (5/6) \oplus (-7/12).$$

Next we define S_{53}. Let Γ_{53} be the lattice $U \oplus A_2 \oplus A_2 \oplus A_3 \oplus A_{11}$, with $G_i (i = 1, 2), H_i (i = 1, 2), J_i (i = 1, 2, 3), \theta_i (1 \le i \le 11)$ as the canonical basis of $A_2 \oplus A_2 \oplus A_3 \oplus A_{11}$, and $\mathcal{O}, F$ as a basis of U as in the case of S_{35}.

Extend Γ_{53} to an index-3 integral over lattice $S_{53} = \Gamma_{53} + \mathbf{Z}s_{53}$, where

$$s_{53} = \mathcal{O} + 2F - (1/3)(2G_1 + G_2 + 2H_1 + H_2) - (2/3)\sum_{i=1}^{11} i\theta_i + \sum_{i=5}^{11} (i-4)\theta_i,$$

$$(\text{set } s = s_{53})\ s^2 = -2, s \cdot F = s \cdot G_1 = s \cdot H_1 = s \cdot \theta_4 = 1,$$

$$s \cdot \mathcal{O} = s \cdot G_2 = s \cdot H_2 = s \cdot J_i = s \cdot \theta_j = 0\ (\forall i; j \neq 4).$$

Moreover, $|\det(S_{53})| = |\det(\Gamma_{53})|/3^2 = 48$.

Note that $\Gamma_{53}^\vee$ is generated by the following, modulo Γ_{53}:

$$h_1 = (1/3)\sum_{i=1}^{2} iG_i, \quad h_2 = (1/3)\sum_{i=1}^{2} iH_i,$$

$$h_3 = (1/4)\sum_{i=1}^{3} iJ_i, \quad h_4 = (1/12)\sum_{i=1}^{11} i\theta_i.$$

$(S_{53})^\vee$ is generated by the following, modulo Γ_{53}:

$$h_3,\ 3h_i,\ h_1 + h_2 + h_4,\ h_1 - h_2,\ h_1 - h_4,\ h_2 - h_4.$$

Noting that $3h_1, 3h_2 \in S_{53}$ and $3h_4 + (h_1 + h_2 + h_4)$ is equal to s (mod Γ_{53}) and hence contained in S_{53}, we see that $(S_{53})^\vee$ is generated by $\varepsilon_1 := h_3, \varepsilon_2 := h_1 - h_4$, modulo S_{53}. As in the case of S_{35}, $(S_{53})^\vee/S_{53}$ is a direct sum of its cyclic subgroups, which are of order 4, 12, and generated by $\varepsilon_1, \varepsilon_2$, modulo S_{53}.

The negative of the discriminant form

$$\begin{aligned}-q_{(S_{53})} &= (-(\varepsilon_1)^2) \oplus (-(\varepsilon_2)^2)\\ &= (3/4) \oplus ((2/3) + (11/12)) = (3/4) \oplus (-5/12).\end{aligned}$$

Claim 3.4. The pair (S_m, T_m) satisfies the conditions of Lemma (3.1) and hence if we let X_m be the unique $K3$ surface with $T_{X_m} = T_m$ then $\operatorname{Pic} X_m = S_m$ (both two equalities here are modulo isometries).

Proof of the claim. We need to show that $q_{T_m} = -q_{S_m}$. Note that $A_n^\vee / A_n = \mathbf{Z}/(n+1)\mathbf{Z}$ and $q_{(A_n)} = (-n/(n+1))$. For the first nine m, if we write $S_m = U \oplus A_{n_1-1} \oplus \cdots A_{n_k-1}$, then

$$q_{S_m} = (-(n_1-1)/n_1) \oplus \cdots \oplus (-(n_k-1)/n_k);$$

moreover, $S_m^\vee/S_m$ is generated by two elements ε_i $(i = 1, 2)$ (ε_i is a simple sum of the natural generators of $S_m^\vee/S_m$) such that for every $a, b \in \mathbf{Z}$ one has $-q_{(S_m)}(a\varepsilon_1 + a\varepsilon_2) = -a^2(\varepsilon_1)^2 - b^2(\varepsilon_2^2)$. For all eleven m, ε_i can be chosen such that $(-\varepsilon_1^2, -\varepsilon_2^2)$ is respectively given as follows:

$$\begin{gathered}(1/2, 17/18), (9/10, 9/10), (1/2, -19/48), (1/2, 121/70),\\ (9/10, 49/30), (-5/6, -17/30), (25/18, 25/18), (-5/6, -17/42),\\ (-11/24, -11/24), (5/6, -7/12), (3/4, -5/12).\end{gathered}$$

On the other hand, $T_m^\vee$ is generated by $(g_1\ g_2) = (e_1\ e_2)T_m^{-1}$, where e_1, e_2 form a canonical basis of T_m which gives rise to the intersection matrix of T_m shown before this claim. Now, the claim follows from the existence of the following isomorphism, which induces $q_{T_m} = -q_{S_m}$:

$$\varphi : T_m^\vee/T_m \to S_m^\vee/S_m; \quad (g_1\ g_2) \mapsto (\varepsilon_1\ \varepsilon_2)B_m.$$

Here B_m is respectively given as:

$$\begin{gathered}\begin{pmatrix}1 & 1\\ 2 & 5\end{pmatrix}, \begin{pmatrix}7 & 0\\ 0 & 7\end{pmatrix}, \begin{pmatrix}0 & 1\\ 11 & 17\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 51\end{pmatrix}, \begin{pmatrix}7 & 0\\ 0 & 17\end{pmatrix}, \begin{pmatrix}-2 & 1\\ 1 & 3\end{pmatrix},\\ \begin{pmatrix}7 & 0\\ 0 & 7\end{pmatrix}, \begin{pmatrix}2 & 3\\ 21 & 10\end{pmatrix}, \begin{pmatrix}2 & 3\\ 3 & -2\end{pmatrix}, \begin{pmatrix}3 & 2\\ 4 & 3\end{pmatrix}, \begin{pmatrix}0 & 1\\ 3 & 4\end{pmatrix}.\end{gathered}$$

Write S_m (resp. Γ_m) as $U \oplus \mathbf{A}(m)$ with $\mathbf{A}(m) = A_{n_1-1} \oplus \cdots \oplus A_{n_k-1}$, for the first nine m (resp. $m = 35, 53$) as in the definitions of them. Let $\mathcal{O}, F$ be a $\mathbf{Z}$-basis of U for all m, as in the definition of S_{35}. By [**PSS**, p. 573, Th. 1], after an (isometric) action of reflections on $S_m = \operatorname{Pic} X_m$, we may assume at the beginning that F is a fiber of an elliptic fibration $f_m : X_m \to \mathbf{P}^1$. Since $\mathcal{O}^2 = -2$, Riemann-Roch Theorem implies that $\mathcal{O}$ is an effective divisor because $\mathcal{O} \cdot F > 0$. Moreover, $\mathcal{O} \cdot F = 1$ implies that $\mathcal{O} = \mathcal{O}_1 + F'$ where $\mathcal{O}_1$ is a cross-section of f_m and F' is an effective divisor contained in fibers. So f_m is a Jacobian elliptic fibration and we can choose $\mathcal{O}_1$ as the zero element of $MW(f_m)$.

Let Λ_m be the lattice generated by all fiber components of f_m. Clearly, $\Lambda_m = \mathbf{Z}F \oplus \Delta, \Delta = \Delta(1) \oplus \cdots \oplus \Delta(r)$ (depending on m), where each $\Delta(i)$ is a negative definite even lattice of Dynkin type A_p, D_q, or E_r, contained in a single reducible singular fiber F_i of f_m and spanned by smooth components of F_i disjoint from $\mathcal{O}_1$.

Claim 3.5. We have:

(1) $\mathrm{Span}_{\mathbf{Z}}\{x \in S_m | x \cdot F = 0, x^2 = -2\} = \Lambda_m = \mathbf{Z}F \oplus \mathbf{A}(m)$; in particular, $r = k$, and there are lattice-isometries: $\Delta \cong \mathbf{A}(m)$ and $\Delta(i) \cong A_{n_i}$ $(i = 1, 2, \ldots, k)$, after relabeling.

(2) There are k singular fibers F_i of type $\widetilde{A}_{n_i - 1}$ $(1 \le i \le k)$ of f_m, and any fiber other than F_i is irreducible.

(3) $MW(f_m) = (0)$ (resp. $\mathbf{Z}/2\mathbf{Z}$, $\mathbf{Z}/3\mathbf{Z}$) for the first nine m (resp. $m = 35, 53$).

Proof. The assertion (2) follows from (1) (see also [**K**, Lemma 2.2]).

The first equality in (1) is clear from Kodaira's classification of elliptic fibers and the Riemann Roch Theorem as used prior to this claim to deduce $\mathcal{O} \ge 0$. The second equality is clear for the cases of the first nine m because then $\mathrm{Pic}\, X_m = S_m = (\mathbf{Z}\mathcal{O} + \mathbf{Z}F) \oplus \mathbf{A}(m)$.

Let $m = 35, 53$. We now show the second equality using Lemma 1.3. Clearly, $\mathbf{Z}F \oplus \mathbf{A}(m)$ is contained in the first term of (1) and hence in Λ_m. One notes that $19 = \text{rank } S_m - 1 \ge \text{rank } \Lambda_m = 1 + \text{ rank } \Delta \ge 1 + \text{ rank } \mathbf{A}(m) = 1 + \sum_{i=1}^{k}(n_i - 1) = 19$. Hence $\Delta = \Delta(1) \oplus \cdots \oplus \Delta(r) \cong \Lambda_m / \mathbf{Z}F$ contains a finite-index sublattice $(\mathbf{Z}F \oplus \mathbf{A}(m))/\mathbf{Z}F \cong \mathbf{A}(m) = A_{n_1 - 1} \oplus \cdots \oplus A_{n_k - 1}$.

Suppose the contrary that the second equality in (1) is not true. Then $\mathbf{A}(m)$ is an index-n $(n > 1)$ sublattice of Δ. By Lemma 1.3, one of Cases (2-1) - (2-3) there occurs.

Case (2-1). Then $m = 35$, f_m has reducible singular fibers of types $\widetilde{A}_1, I_{13}^*$ and no other reducible fibers. This leads to that $72 = |\mathrm{Pic}\, X_m| = (2 \times 4)/|MW(f_m)|^2$, a contradiction (cf. [**S**]).

Case (2-2). Then $m = 53$, f_m has reducible singular fibers of types $\widetilde{A}_2, I_{12}^*$ and no other reducible fibers. This leads to that $48 = |\mathrm{Pic}\, X_m| = (3 \times 4)/|MW(f_m)|^2$, a contradiction.

Case (2-3). Then $m = 35$, f_m has reducible singular fibers of types $\widetilde{A}_1, I_{12}, IV^*$ and no other reducible fibers. Since $72 = |\mathrm{Pic}\, X_m| = (2 \times 12 \times 3)/|MW(f_m)|^2$, we have $MW(f_m) = (0)$ and $S_m = \mathrm{Pic}\, X_m = \mathbf{Z}\mathcal{O}_1 + \Lambda_m = \mathbf{Z}\mathcal{O}_1 + (\ \mathbf{Z}F \oplus \Delta) = \mathbf{Z}\mathcal{O}_1 + (\ \mathbf{Z}F \oplus A_1 \oplus A_{11} \oplus E_6)$.

By the Riemann-Roch theorem and the fact that $(s_m)^2 = -2$, $s_m.F = 1$ and $MW(f_m) = (0)$, we see that $s_m = \mathcal{O}_1 \pmod{\Lambda_m}$. This, together with the fact that $\mathcal{O} = \mathcal{O}_1 \pmod{\Lambda_m}$ and the definition of s_m, implies that $(1/2)(G + H + D) \in \Lambda_m$, where $D = \sum_{i=1}^{6} i\theta_i + \sum_{i=7}^{11}(12 - i)\theta_i$.

Consider the index-2 extension

$$A_1 \oplus A_{11} \oplus (A_1 \oplus A_5) = \mathbf{A}(m) \cong (\mathbf{Z}F \oplus \mathbf{A}(m))/\mathbf{Z}F \subseteq (\mathbf{Z}F \oplus \Delta)/\mathbf{Z}F \\ \cong \Delta = A_1 \oplus A_{11} \oplus E_6.$$

The proof of Lemma 1.3 shows that (the first summand A_1 in this rearranged $\mathbf{A}(m)$) $\oplus\mathbf{Z}F =$ (the summand A_1 in Δ) $\oplus\mathbf{Z}F$, (the summand A_{11} in $\mathbf{A}(m)$) $\oplus\mathbf{Z}F =$ (the summand A_{11} in Δ) $\oplus\mathbf{Z}F$, and (the summand $(A_1 \oplus A_5)$ in $\mathbf{A}(m)$) $\oplus\mathbf{Z}F \subseteq$ (the summand E_6 in Δ) $\oplus\mathbf{Z}F$. So we may assume that, mod $\mathbf{Z}F$, G is the $\mathbf{Z}$-generator of the first summand A_1 in Δ, θ_i $(1 \le i \le 11)$ form a $\mathbf{Z}$-basis of the summand A_{11} in Δ, and H is contained in the summand E_6 in Δ.

In particular, for $(G+H+D)/2 \in \Lambda_m = \mathbf{Z}F \oplus \Delta = \mathbf{Z}F \oplus (A_1 \oplus A_{11} \oplus E_6)$, we have, mod $\mathbf{Z}F$, $G/2 \in A_1$, $H/2 \in E_6$, and $D/2 \in A_{11}$. We reach a contradiction to the above observation that the A_1 in Δ is generated by G over $\mathbf{Z}$.

Therefore, the second equality of (1) is true. So there is an isometry $\Phi : \Delta \cong \Lambda_m/\mathbf{Z}F \cong \mathbf{A}(m)$. Now the rest of (1) follows from Lemma 1.3.

The assertion (3) follows from the fact in [**S**, Th. 1.3], that $MW(f_m)$ is isomorphic to the factor group of $\operatorname{Pic} X_m$ modulo $(\mathbf{Z}\mathcal{O}_1 + \mathbf{Z}F) \oplus \Delta$, where the latter is equal to $(\mathbf{Z}\mathcal{O} + \mathbf{Z}F) + \Delta = (\mathbf{Z}\mathcal{O} + \mathbf{Z}F) \oplus \mathbf{A}(m) = U \oplus \mathbf{A}(m)$. This proves the claim.

The existence of singular fibers F_i $(i = 1, 2, \ldots, k)$ of type I_{n_i}, the fact that the sum of Euler numbers of singular fibers of f_m is 24, the fact that each fiber other than F_i is irreducible, and [**MP3**, Lemma 3.1 and Proposition 3.4] imply that f_m is semi-stable. Hence F_i $(i = 1, 2, \ldots, k)$ is of type I_{n_i}, there are $\chi(X_m) - \sum_i (n_i - 1) - k = 6 - k$ fibers of type I_1, and f_m is of type $[1, 1, \ldots, 1, n_1, \ldots, n_k]$, i.e., of type m after an easy case-by-case check. Moreover, $(m, MW(f_m)) = (m, G_m)$ for all eleven m by the last claim. □

Remark 3.6. We note that $S_{35} = U \oplus A_1 \oplus A_{11} \oplus E_6$. This is because the lattices T_{35} and the one on the right hand side satisfy all conditions of Lemma 3.1 by an easy check. In particular, using [**MP3**, Lemma 3.1 and Proposition 3.4] as in the proof of Lemma 3.3, we can show that there is a Jacobian elliptic fibration $\tau_m : X_m \to \mathbf{P}^1$ $(m = 35)$ with singular fibers $I_1, I_1, I_2, I_{12}, IV^*$ and with $MW(\tau_m) = (0)$.

Lemma 3.7. *Let $f : X \to \mathbf{P}^1$ be of type m as in Theorem* 0.2. *Then we have:*

(1) *If $m = 48$, then $MW(f) \neq \mathbf{Z}/2\mathbf{Z}$, or $\mathbf{Z}/4\mathbf{Z}$.*
(2) *If $m = 4$, then $MW(f) \neq \mathbf{Z}/2\mathbf{Z}$.*
(3) *If $m = 31$, then $MW(f) \neq \mathbf{Z}/2\mathbf{Z}$.*
(4) *If $m = 44$, then $MW(f) \neq \mathbf{Z}/2\mathbf{Z}$.*
(5) *If $m = 69$, then it is impossible that $MW(f)$ is $\mathbf{Z}/2\mathbf{Z}$ with $s = (0, 0, 0, 0, 2, 6)$ as its generator* (*see Remark* 0.4).

(6) *If* $m = 92$, *then* $MW(f) \neq \mathbf{Z}/2\mathbf{Z}$.

Proof. Let $f : X \to \mathbf{P}^1$ be of type m as in Theorem 0.2.

(1) Assume that f is of type $m = 48$ and $MW(f) \supseteq \mathbf{Z}/2\mathbf{Z}$. We will show that $MW(f) \supseteq \mathbf{Z}/8\mathbf{Z}$ which will imply (1).

$m = 48$ means that the singular fiber type of f is $I_1, I_1, I_2, I_4, I_8, I_8$. Using the height pairing in [**S**] or the Table in [**MP3**], we may assume that $MW(f)$ contains $s = (0,0,0,0,4,4)$ as a 2-torsion section after suitable labeling of fiber components.

Let Y, a $K3$ surface again, be the minimal resolution of the quotient surface $X/\langle s\rangle$. f on X induces a Jacobian semi-stable elliptic fibration $g : Y \to \mathbf{P}^1$ of singular fiber type $I_2, I_2, I_4, I_8, I_4, I_4$ where these 6 ordered singular fibers are respectively "images" of ordered singular fibers on X.

To be precise, let $\sigma : \widetilde{X} \to X$ be the blowing-up of all 8 intersections in the first 4 singular fibers of f of types I_1, I_1, I_2, I_4. Then $Y = \widetilde{X}/\langle s\rangle$ and the $\mathbf{Z}/2\mathbf{Z}$-covering $\pi : \widetilde{X} \to Y$ is branched along 4 disjoint curves $\theta_j^{(i)}$, where $(i,j) = (1,1),(2,1),(3,1),(3,3),(4,1),(4,3), (4,5),(4,7)$. Here we choose the common image of the zero section and the 2-torsion section s of f, as the zero section O_1 of g, and label clock or anti-clockwise the i-th singular fiber of g of type I_{n_i} as $\sum_{j=0}^{n_i-1} \theta_j^{(i)}$ so that O_1 passes through $\theta_0^{(i)}$, where $[n_1, \dots, n_6] = [2,2,4,8,4,4]$.

Note that (Y,g) is of type $m = 103$ in the Table of [**MP3**] and hence there is a 4-torsion section t of g equal to $(0,0,2,2,1,1)$ or $(0,0,1,2,1,2)$ or $(0,0,1,2,2,1)$, after choosing either clockwise or counterclockwise labeling of fiber components, where for orders of six fibers of g we use the current indexing inherited from that of f.

If $t = (0,0,1,2,1,2)$ or $(0,0,1,2,2,1)$, then t meets the branch locus of π transversally at one point only so that $\pi^{-1}(t)$ is a smooth irreducible curve and $\pi : \pi^{-1}(t) \to t$ is a double cover with exactly one ramification point, a contradiction to Hurwitz's genus formula applied to the covering map π.

Thus $t = (0,0,2,2,1,1)$. A check using height pairing shows that $\pi^{-1}(t)$ is a disjoint union of two 8-torsion sections of f. Hence $MW(f) \supseteq \mathbf{Z}/8\mathbf{Z}$. Indeed, $MW(f) = \mathbf{Z}/8\mathbf{Z}$ by [**MP3**]. This proves (1).

Now assume that f is of type $m = 4$ (resp. $m = 31$, $m = 44$, $m = 69$ with $MW(f) = \langle s = (0,0,0,0,2,6)\rangle$, or $m = 92$) and $MW(f) \supseteq \mathbf{Z}/2\mathbf{Z}$. Then $MW(f)$ contains a unique 2-torsion section $s = (0,0,0,0,0,8)$ (resp. $s = (0,0,0,0,0,8)$, $s = (0,0,0,0,2,6)$, $s = (0,0,0,0,2,6)$, $s = (0,0,0,2,2,4)$) (cf. [**MP3**]). As in (1) we can show that f induces a Jacobian semi-stable elliptic fibration g on the minimal resolution Y of $X/\langle s\rangle$. The singular fiber type of g is $I_{n_1} + \cdots + I_{n_6}$ where $[n_1, \dots, n_6]$ is equal to $[2,2,2,2,8,8]$ (resp. $[2,2,4,4,4,8]$, $[2,2,4,8,2,6]$, $[2,4,4,6,2,6]$, $[2,6,8,2,2,4]$) and hence is of type $m = 94$ (resp. $m = 103$, $m = 97$, $m = 104$, or $m = 97$) in the Table of [**MP3**]. Now the inverse on X of the 2-torsion section $t = (0, 0, 0, 0, 4, 4)$ (resp. $t = (0, 0, 0, 2, 2, 4)$, $t = (0, 0, 0, 4, 1, 3)$, t is one of $(0, 2, 2, 0, 1, 3)$ and $(1, 2, 2, 3, 0, 0)$, or $t = (0, 0, 0, 4, 1, 2)$) on Y is a disjoint union of two 4-torsion sections of f. Hence $MW(f) \supseteq \mathbf{Z}/4\mathbf{Z}$. Indeed, $MW(f) = \mathbf{Z}/4\mathbf{Z}$ by [**MP3**]. This proves (2)-(6). The proof of the lemma is completed. □

Lemma 3.8. *Let $f : X \to \mathbf{P}^1$ be of type m as in Theorem* 0.2. *Then each of the following pairs $(m, MW(f))$ occurs:*

$$(69, \mathbf{Z}/2\mathbf{Z} = \langle(0, 1, 1, 0, 0, 6)\rangle), (69, \mathbf{Z}/4\mathbf{Z}), (92, \mathbf{Z}/4\mathbf{Z}),$$
$$(32, \mathbf{Z}/3\mathbf{Z}), (37, \mathbf{Z}/3\mathbf{Z}), (44, \mathbf{Z}/4\mathbf{Z}), (55, \mathbf{Z}/2\mathbf{Z}).$$

Proof. The idea of the proof for the existence of the pair $(m, MW(f)) = (69, \mathbf{Z}/4\mathbf{Z})$ is as follows. By [**MP3**], $s = (0, 1, 1, 0, 1, 3)$ is the generator of $MW(f) = \mathbf{Z}/4\mathbf{Z}$. As in the proof of Lemma 3.7, the minimal resolution Y of $X/\langle 2s\rangle$ is of type $m = 104$. The detailed proof of the existence is given below.

Let $g : Y \to \mathbf{P}^1$ be of type $m = 104$. By the Table in [**MP3**], $MW(g) = \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$ and we may assume that g has singular fibres $\sum_{j=0}^{n_i-1} \theta(i)_j$ $(i = 1, \dots, 6)$ of type I_{n_i}, and two 2-torsion sections $t_1 = (0, 2, 2, 0, 1, 3), t_2 = (1, 2, 2, 3, 0, 0)$, after suitably indexing singular fibers so that $[n_1, \dots, n_6] = [2, 4, 4, 6, 2, 6]$. It is easy to check the following relation (cf. [**S**, Lemma 8.1] or [**M**, Formula (2.5)]), where O_1, F are the zero section and a general fiber of g,

$$2t_2 \sim 2(O_1 + 2F) - (\theta(1)_1 + \theta(2)_1 + 2\theta(2)_2 + \theta(2)_3 + \theta(3)_1 + 2\theta(3)_2 + \theta(3)_3 + \theta(4)_1 + 2\theta(4)_2 + 3\theta(4)_3 + 2\theta(4)_4 + \theta(4)_5).$$

Hence we get a relation

$$D = \theta(1)_1 + \theta(2)_1 + \theta(2)_3 + \theta(3)_1 + \theta(3)_3 + \theta(4)_1 + \theta(4)_3 + +\theta(4)_5 \sim 2L$$

for some integral divisor L. Let $\pi : \widetilde{X} \to Y$ be the $\mathbf{Z}/2\mathbf{Z}$-cover, branched along D and induced from the above relation. Then g induces an elliptic fibration $f : \widetilde{X} \to \mathbf{P}^1$ so that the relatively minimal model (X, f) of $(\widetilde{X}, f)$ is of type $m = 69$. The inverse on X of O_1 is a disjoint union of two sections, one of which will be fixed as O of f. Now the inverse on X of the 2-torsion section t_1 on Y is a disjoint union of two 4-torsion sections of f. Hence $MW(f) = \mathbf{Z}/4\mathbf{Z}$ by the Table in [**MP3**]. This proves the existence of the pair $(m, MW(f)) = (69, \mathbf{Z}/4\mathbf{Z})$.

The existence of other pairs is similar. Here we just show which Y, t_1, t_2 we should choose. To be precise, we let $g : Y \to \mathbf{P}^1$ be of type $m = 52$ (resp. $m = 97$; $m = 91$; $m = 110$; $m = 97$; $m = 104$) with singular fibers of type $I_{n_1} + \cdots + I_{n_6}$ with $[n_1, \dots, n_6] = [2, 1, 1, 6, 8, 6]$ (resp. $[2, 6, 8, 2, 2, 4]$; $[3, 3, 6, 6, 1, 5]$; $[3, 3, 6, 6, 3, 3]$; $[2, 2, 4, 8, 2, 6]$; $[2, 2, 6, 6, 4, 4]$) and we let

$t_1 = O_1$ be the zero section and $t_2 = (1, 0, 0, 3, 4, 0)$ the 2-torsion section (resp. $t_1 = (0, 0, 4, 1, 1, 2)$ and $t_2 = (1, 3, 4, 0, 0, 0)$ two 2-torsion sections; $t_1 = O_1$ and $t_2 = (1, 1, 2, 2, 0, 0)$ a 3-torsion section; $t_1 = O_1$ and $t_2 = (1, 1, 2, 2, 0, 0)$ a 3-torsion section; $t_1 = (0, 0, 0, 4, 1, 3)$ and $t_2 = (1, 1, 2, 4, 0, 0)$ two 2-torsion sections; $t_1 = O_1$ and $t_2 = (1, 1, 3, 3, 0, 0)$ a 2-torsion section). Then as in the above paragraph, the minimal model X of a $\mathbf{Z}/n\mathbf{Z}$-cover with $n = 2$ (resp. $n = 2$; $n = 3$; $n = 3$; $n = 2$; $n = 2$) of Y has an elliptic fibration $f : X \to \mathbf{P}^1$, induced from g, of type $m = 69$ (resp. $m = 92$; $m = 32$; $m = 37$; $m = 44$; $m = 55$) such that the inverse on X of t_1 is a disjoint union of O and $s = (0, 1, 1, 0, 0, 6)$ (resp. a disjoint union of two 4-torsion sections; a disjoint union of O and two 3-torsion sections; a disjoint union of O and two 3-torsion sections; a disjoint union of two 4-torsion sections; a disjoint union of O and a 2-torsion section), whence $MW(f)$ is equal to $\mathbf{Z}/2\mathbf{Z} = \{O, s\}$ (resp. $\mathbf{Z}/4\mathbf{Z}$; $\mathbf{Z}/3\mathbf{Z}$; $\mathbf{Z}/3\mathbf{Z}$; $\mathbf{Z}/4\mathbf{Z}$; $\mathbf{Z}/2\mathbf{Z}$) by the Table in [**MP3**].

This completes the proof of the lemma and also that of Theorem 0.2. □

4. Uniqueness for extremal elliptic $K3$ surfaces with large torsion groups.

The goal of this section is to prove Theorem 0.3.

In the case where $MW(f) \supseteq \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$, namely, $m = 94, 97, 98, 103, 104, 112$, the uniqueness problem has already been considered in §7 [**MP3**] by using double sextics, and they are all unique. Hence we need to prove the cases when $MW(f) \cong \mathbf{Z}/4\mathbf{Z}$, $\mathbf{Z}/5\mathbf{Z}$, $\mathbf{Z}/6\mathbf{Z}$, $\mathbf{Z}/7\mathbf{Z}$, $\mathbf{Z}/3\mathbf{Z} \times \mathbf{Z}/3\mathbf{Z}$.

As we have seen in §1, if $MW(f)$ has an element of order $N \geq 3$, then $f : X \to \mathbf{P}^1$ is obtained as the pull-back surface of the elliptic surface, $\psi_{1,N} : E_1(N) \to X_1(N)$, by some morphism $g : \mathbf{P}^1 \to X_1(N)$. Since $X_1(N)$ should be isomorphic to $\mathbf{P}^1$ and X is a $K3$ surface in our case, $N \leq 8$ by [**C**]. Thus our proof of Theorem 0.3 is reduced to showing the uniqueness of g up to $\mathrm{Aut}(\mathbf{P}^1)$ for each case. Hence it is enough to prove the following:

Proposition 4.1. *Let $g : \mathbf{P}^1 \to X_1(N)$ be the morphism as above. Then g is unique except $m = 49$.*

By comparing the degree of the j-functions, we can easily check the following table:

$MW(f)$	$\deg g$	m
$\mathbf{Z}/4\mathbf{Z}$	4	$4, 31, 44, 69, 92$
$\mathbf{Z}/5\mathbf{Z}$	2	$9, 49, 105$
$\mathbf{Z}/6\mathbf{Z}$	2	$35, 53, 63, 95, 108$
$\mathbf{Z}/7\mathbf{Z}$	1	30
$\mathbf{Z}/3\mathbf{Z} \times \mathbf{Z}/3\mathbf{Z}$	2	110

Table 4.2.

One can see that the uniqueness for the case $MW(f) \cong \mathbf{Z}/7\mathbf{Z}$ ($m = 30$) immediately from the table.

Let us consider the cases of $\deg g = 2$. Our goal is to show that g is unique up to $\mathrm{Aut}(X_1(N))(\cong \mathrm{Aut}(\mathbf{P}^1))$ except $m = 49$.

Case $m = 9$**.** $f : X \to \mathbf{P}^1$ has two I_{10} fibers. This means that the branch points of g are 2 points over which $\psi_{1,5}$ has I_5 fibers. The choice of such two points is unique and g is determined by the branch points. Hence g is unique.

For cases $m = 35, 53, 63, 95, 105, 108$, we can prove the uniqueness in a similar way to that for $m = 9$. Hence we omit it.

Case $m = 110$**.** In this case, $f : X \to \mathbf{P}^1$ is obtained as the pull-back surface of $\psi_{3,3} : E_3(3) \to X_3(3)$ by a degree 2 map $g : \mathbf{P}^1 \to X_3(3)$. $\psi_{3,3}$ has 4 singular fibers, all of which are of type I_3. By [**MP1**, Table 5.3], $E_3(3)$ is given by the Weierstrass equation as follows:

$$y^2 = x^3 + (-3s^2 + 24s)x + (2s^6 + 40s^3 - 16),$$

where s is an inhomogeneous coordinate of $X_3(3) \cong \mathbf{P}^1$. The four I_3 fibers are over $-1, -\omega, \omega^2$ and ∞, where $\omega = \exp(2\pi\sqrt{-1}/3)$.

Consider two fiber preserving automorphisms of $E_3(3)$:

$$\tau_1 : (x, y, s) \mapsto \left(\frac{-3}{(s+1)^2}x, \frac{3\sqrt{-3}}{(s+1)^3}y, \frac{-s+2}{s+1} \right),$$

and

$$\tau_2 : (x, y, s) \mapsto (\omega x, y, \omega s).$$

These automorphisms induce permutations of the I_3 fibers. Since X is a double covering of $E_3(3)$, it is uniquely determined by the branch locus which is two I_3 fibers. Therefore, using τ_1 and τ_2, we can show that $f : X \to \mathbf{P}^1$ is unique.

Putting the case $m = 49$ the aside, we consider the cases of $\deg g = 4$. There are 5 cases: $m = 4, 31, 44, 69, 92$.

The degree of the j-invariant of $E_1(4)$ is 6, as it has three singular fibers I_1^*, I_4 and I_1. With a suitable affine coordinate of $X_1(4)$, we may assume that these singular fibers are over 0, 1 and ∞, respectively. Since the degree of the j-invariant of $f : X \to \mathbf{P}^1$ is 24, the degree of g is 4, and is branched only at 0, 1 and ∞. By [**MP1**, Table 7.1] and the Riemann-Hurwitz formula for $g : \mathbf{P}^1 \to X_1(4)$, we have the following table on the ramification types over each branch point.

m	The ramification types over 0, 1 and ∞
4	(4), (4), $(1,1,1,1)$
31	$(2,2)$, (4), $(2,1,1)$
44	(4), $(3,1)$, $(2,1,1)$
69	$(2,2)$, $(3,1)$, $(3,1)$
92	(4), $(2,1,1)$, $(3,1)$

Table 4.3.

Here the notation $(e_1, \dots, e_k)$ means that $g^{-1}(p)$ ($p \in \{0, 1, \infty\}$) consists of k points, $q_1, \dots, q_k$, and the ramification index at q_j is e_j.

To show the uniqueness, it is enough to show that g assigned with the ramification types as above is unique up to covering isomorphisms over $X_1(4)$. Let us start with the following lemma.

Lemma 4.4. *Let $g : \mathbf{P}^1 \to X_1(4)$ be one of the degree 4 maps in Table 4.3. Let $\alpha : C \to \mathbf{P}^1$ be the Galois closure, and put $\hat{g} = g \circ \alpha$. Then we have the following:*

$m = 4$: $g = \hat{g}$ and g is a 4-fold cyclic covering.
$m = 31$: $\deg \hat{g} = 8$, $C \cong \mathbf{P}^1$ and $\mathrm{Gal}(\hat{g}) \cong \mathcal{D}_8$.
$m = 44, 92$: $\deg \hat{g} = 24$, $C \cong \mathbf{P}^1$ and $\mathrm{Gal}(\hat{g}) \cong \mathcal{S}_4$.
$m = 69$: $\deg \hat{g} = 12$, $C \cong \mathbf{P}^1$ and $\mathrm{Gal}(\hat{g}) \cong \mathcal{A}_4$.

Proof. The monodromy around the branch points gives a permutation representation of $\pi_1(\mathbf{P}^1 \setminus \{0, 1, \infty\})$ to $\mathcal{S}_4$; the basic loops γ_0, γ_1 and γ_∞ about 0, 1 and ∞, respectively map to permutations σ_0, σ_1 and σ_∞. The cycle structure of each permutation is the same as the ramification type over the corresponding point. These permutations satisfy the identity $\sigma_0\sigma_1\sigma_\infty = 1$ in $\mathcal{S}_4$ and generate a transitive subgroup, G, in $\mathcal{S}_4$. Note that this G is nothing but the Galois group of $\hat{g} : C \to X_1(4)$. We apply this argument to each case, and obtain the following table:

m	The cycle structure of σ_0, σ_1 and σ_∞	G
4	(4), (4), (1, 1, 1, 1)	$\mathbf{Z}/4\mathbf{Z}$
31	(2, 2), (4), (2, 1, 1)	$\mathcal{D}_8$
44	(4), (3, 1), (2, 1, 1)	$\mathcal{S}_4$
69	(2, 2), (3, 1), (3, 1)	$\mathcal{A}_4$
92	(4), (2, 1, 1), (3, 1)	$\mathcal{S}_4$

Table 4.5.

Now all we need to show Is: $C \cong \mathbf{P}^1$. Our argument is based on the following elementary fact:

Fact 4.6. Let x be a point on C, and put $G_x = \{\tau \in G | \tau(x) = x\}$. Then

G	The order of G_x
$\mathbf{Z}/4\mathbf{Z}$	1, 2, 3
$\mathcal{S}_4$	1, 2, 3, 4
$\mathcal{A}_4$	1, 2, 3
$\mathcal{D}_8$	1, 2, 4

We prove $C \cong \mathbf{P}^1$ case by case.

Case $m = 4$**.** As $G = \mathbf{Z}/4\mathbf{Z}$, $\deg \hat{g} = \deg g$, and α is the identity.

Case $m = 31$**.** Since $G = \mathcal{D}_8$, $\deg \alpha = 2$. Let ι be an element of order 2 such that $C/\langle\iota\rangle \cong \mathbf{P}^1$. As g is not Galois, ι is not contained in the center of $\mathcal{D}_8$. If α is branched over $g^{-1}(0)$, then $\hat{g}^{-1}(0)$ consists of two points, each of which has the ramification index 4. This means that ι belongs to the center of $\mathcal{D}_8$, which leads us to a contradiction. Hence the branch points of α are two points in $g^{-1}(\infty)$ which are unramified points of g. Hence $C \cong \mathbf{P}^1$.

Cases $m = 44$, 92**.** By Fact 4.6 and $\mathrm{Gal}(C/\mathbf{P}^1) \cong \mathcal{S}_4$, points over 0, 1 and ∞ have the ramification indices 4, 3 and 2, respectively. By the Riemann-Hurwitz formula, we have $C \cong \mathbf{P}^1$.

Case $m = 69$**.** By Fact 4.6, points over 0, 1 and ∞ have the ramification indices 2, 3 and 3, respectively. By the Riemann-Hurwitz formula, $C \cong \mathbf{P}^1$.

This completes our proof for Lemma 4.4. □

The following classical fact is a key to prove Theorem 0.3 in the case where $MW(f) \cong \mathbf{Z}/4\mathbf{Z}$.

Fact 4.7 ([**Na**, pp. 31-32])**.** For a suitable choice of an affine coordinate, w and z, of $X_1(4)$ and $\mathbf{P}^1$, respectively, the map in Table 4.5 can be given by the rational functions as follows:

$$w = z^4 \qquad m = 4$$

$$w = -\frac{(z^4-1)^2}{4z^2} \qquad m = 31$$

$$w = \left(\frac{z^4 + 2\sqrt{3}z^2 - 1}{z^4 - 2\sqrt{3}z^2 - 1}\right)^3 \qquad m = 69$$

$$w = \frac{(z^8 + 14z^4 + 1)^3}{108z^4(z^4-1)^4} \qquad m = 44, 92.$$

Fact 4.7 implies that the Galois coverings described in Lemma 4.4 are essentially unique up to isomorphisms over $\mathbf{P}^1$. The morphisms g in Lemma 4.4 are corresponding to a subgroup of index 4 of G, and for each case, such subgroups are conjugate to each other. This shows that the pull-back morphisms, g, are unique up to covering isomorphisms over $X_1(4)$. Therefore we have Proposition 4.1 in the case where $MW(f) \cong \mathbf{Z}/4\mathbf{Z}$.

Remark 4.8. We can prove the uniqueness for $m = 94, 98, 103, 112$ in a similar way to the case $MW(f) \cong \mathbf{Z}/4\mathbf{Z}$.

We now go on to show that the uniqueness does not hold for $m = 49$.

For the case $m = 49$, as we have seen before, $f :\rightarrow \mathbf{P}^1$ is obtained as the pull-back surface of $\psi_{1,5} : E_1(5) \rightarrow X_1(5)$ by a degree 2 map $g : \mathbf{P}^1 \rightarrow X_1(5)$. $\psi_{1,5}$ has 4 singular fibers. By [**MP1**, Table 5.3], $E_1(5)$ is given by the following Weierstrass equation:

$$y^2 = x^3 - 3(s^4 - 12s^3 + 14s^2 + 12s + 1)x + 2(s^6 - 18s^5 + 75s^4 + 75s^2 + 18s + 1),$$

where s is an inhomogeneous coordinate of $X_1(5) \cong \mathbf{P}^1$. The two I_5 fibers are over $s = 1$ and $s = \infty$, and the two I_1 fibers are over $s = (11 \pm 5\sqrt{5})/2$.

For $m = 49$, There are 4 possible cases for the pull-back morphism depending on the branch points as follows:

	The branch points of g
(1)	0 and $(11+5\sqrt{5})/2$
(2)	0 and $(11-5\sqrt{5})/2$
(3)	∞ and $(11+5\sqrt{5})/2$
(4)	∞ and $(11-5\sqrt{5})/2$

We denote the pull-back morphisms by g_i ($i = 1, 2, 3, 4$) corresponding to the cases as above, and let $f_i :\to \mathbf{P}^1$ denote the pull-back surface by g_i. Then we have:

Proposition 4.9. *There exists φ in Question* 0.1 *between either X_1 and X_4 or X_2 and X_3, while there is no such φ between the two pull-back surfaces for other combinations.*

Proof. Consider an automorphism, τ, of $E_1(5) \to X_1(5)$ given by

$$\tau : (x, y, s) \mapsto \left(\frac{1}{s^2}x, \frac{1}{s^3}y, -\frac{1}{s}\right).$$

With τ, the points 0 and $(11+5\sqrt{5})/2$ map to ∞ and $(11-5\sqrt{5})/2$, respectively. Our first assertion follows from this fact. For the second, by using τ, it is enough to show that there is no φ in Question 0.1 between the pull-back surfaces X_1 and X_2.

Suppose that there exists $\varphi : X_1 \to X_2$ as Question 0.1. Then we have:

Claim 4.10. φ induces an automorphism $\hat{\varphi} : X_1(5) \to X_1(5)$ such that $0 \mapsto \infty$, $\infty \mapsto 0$, $(11+5\sqrt{5})/2 \mapsto (11-5\sqrt{5})/2$, and $(11-5\sqrt{5})/2 \mapsto (11+5\sqrt{5})/2$.

Since there is no fractional linear transformation as above, the second assertion follows.

Proof of the Claim. Let ι_i ($i = 1, 2$) be fiber preserving involutions on X_i ($i = 1, 2$) determined by the pull-back morphisms g_i. Let $\overline{\varphi}$ and $\bar{\iota}_i$ ($i = 1, 2$) be the restrictions of each morphism to the zero sections of X_1 and X_2. $\varphi^{-1} \circ \iota_2 \circ \varphi$ gives rise to another fiber preserving involution on X_1. Under $\varphi^{-1} \circ \iota_2 \circ \varphi$, I_{10}, I_5, I_2 fibers map to I_{10}, I_5, I_2 fibers, respectively. Hence $\overline{\varphi}^{-1} \circ \bar{\iota}_2 \circ \overline{\varphi} = \bar{\iota}_1$ or id, but the latter case does not occur since $\bar{\iota}_2 \neq id$. Thus we have an isomorphism $\hat{\varphi} : X_1(5) \to X_1(5)$, and it is easy to see that $\widetilde{\varphi}$ has the desired property.

This finishes the proof of Proposition 4.1.

References

[AT] E. Artal Bartolo and H. Tokunaga, *Zariski pairs of index* 19 *and Mordell-Weil groups of* $K3$ *surfaces*, Proc. London. Math. Soc., III, Ser. **80**(**1**) (2001), 127-144, MR 2001a:14024.

[BPV] W.P. Barth, C.A.M. Peters and A.J.H.M. Van de Ven, *Compact Complex Surfaces*, Springer, Berlin, 1984, MR 86c:32026, Zbl 718.14023.

[C] D. Cox, *Mordell-Weil groups of elliptic curves over* $\mathbf{C}(t)$ *with* $p_g = 0$ *or* 1, Duke Math. Journal, **49** (1982), 677-689, MR 84d:14029, Zbl 503.14018.

[CP] D. Cox and W. Parry, *Torsion in elliptic curves over* $k(t)$, Compos. Math., **41** (1980), 337-354, MR 81k:14035, Zbl 442.14015.

[Ko] K. Kodaira, *On compact analytic surfaces* II, Ann. of Math., **77** (1963), 563-626, MR 32 #1730, Zbl 118.15802.

[K] S. Kondo, *Automorphisms of algebraic* $K3$ *surfaces which act trivially on Picard groups*, J. Math. Soc. Japan, **44** (1992), 75-98, MR 93e:14046, Zbl 763.14021.

[M] R. Miranda, *Component numbers for torsion sections of semistable elliptic surfaces*, Classification of Algebraic Varieties (L'Aquila, 1992), Contemporary Math., **162** (1994), 293-311, Amer. Math. Soc., Providence, RI, MR 95d:14033, Zbl 839.14041.

[MP1] R. Miranda and U. Persson, *On extremal rational elliptic surfaces*, Math. Z., **193** (1986), 537-558, MR 88a:14044, Zbl 652.14003.

[MP2] ———, *Configurations of* I_n *fibers on elliptic* $K3$ *surfaces*, Math. Z., **201** (1989), 339-361, MR 90h:14051, Zbl 694.14019.

[MP3] ———, *Mordell-Weil Groups of extremal elliptic* $K3$ *surfaces*, in 'Problems in the theory of surfaces and their classification' (Cortona, 1988), Symposia Mathematica, **XXXII** (1991), 167-192, Academic Press, London, MR 95f:14071, Zbl 827.14013.

[Mo] D.R. Morrison, *On* $K3$ *surfaces with large Picard number*, Inv. Math., **75** (1984), 105-121, MR 85j:14071, Zbl 526.14024.

[Na] M. Namba, *Branched coverings and algebraic functions*, **161**, Pitman Research Note in Math., Longman Scientific & Technical, Harlow, 1987, MR 89d:14020, Zbl 706.14017.

[Ni] V.V. Nikulin, *Integral symmetric bilinear forms and some of their applications*, Math. USSR Izv., **14** (1980), 103-167, Zbl 427.10014.

[P] U. Persson, *Double sextics and singular* $K3$ *surfaces*, in 'Algebraic geometry, Sitges 1983. Proceedings', Lecture Notes in Mathematics, **1124**, Springer Verlag, Berlin, Heidelberg, New York, Tokyo, (1985), 262-328, MR 87i:14036, Zbl 603.14023.

[PSS] I.I. Pjateckiĭ-Šapiro and I.R. Šafarevič, *Torelli's theorem for algebraic surfaces of type* $K3$, Math. USSR Izv., **5** (1971), 547-588, MR 44 #1666, Zbl 253.14006.

[Se] J.P. Serre, *A course in arithmetic*, Graduate Texts in Mathematics, **7** Springer-Verlag, New York-Heidelberg, 1973, MR 49 #8956, Zbl 256.12001.

[SZ] I. Shimada and D.-Q. Zhang, *Classification of extremal elliptic* $K3$ *surfaces and fundamental groups of open* $K3$ *surfaces*, Nagoya Math. J., **161** (2001), 23-54, CMP 1 820 211.

[S] T. Shioda, *On the Mordell-Weil lattices*, Comment. Math. Univ. St. Paul., **39** (1990), 211-240, MR 91m:14056, Zbl 725.14017.

[SI] T. Shioda and H. Inose, *On singular $K3$ surfaces*, Complex Analysis and Algebraic Geometry: Papers Dedicated to K. Kodaira, 119-136 Iwanami Shoten and Cambridge University Press, London, 1977, MR 56 #371, Zbl 374.14006.

[Y] Q. Ye, *On extremal elliptic $K3$ surfaces*, J. Korean Math. Soc., **36**(**6**) (1999), 1091-1113, MR 2001g:14062.

Received February 15, 2000. The first author was partially supported by DGES PB97-0284-C02-02. The second author's research was partly supported by the Grant-in-Aid for Encouragement of Young Scientists 09740031 from the Ministry of Education, Science and Culture. The third author received financial support by the JSPS-NUS program.

Departamento de Matemáticas
Universidad de Zaragoza
Campus Plaza San Francisco s/n
E-50009 Zaragoza, Spain
E-mail address: artal@posta.unizar.es

Department of Mathematics
Kochi University
Kochi 780, Japan
E-mail address: tokunaga@math.kochi-u.ac.jp

Department of Mathematics
National University of Singapore
Lower Kent Ridge Road
Singapore 119260
E-mail address: matzdq@math.nus.edu.sg

PACIFIC JOURNAL OF MATHEMATICS
Vol. 202, No. 1, 2002

QUASIREGULAR MAPPINGS AND $\mathcal{WT}$-CLASSES OF DIFFERENTIAL FORMS ON RIEMANNIAN MANIFOLDS

D. Franke, O. Martio, V.M. Miklyukov, M. Vuorinen, and R. Wisk

The purpose of this paper is to study the relations between quasiregular mappings on Riemannian manifolds and differential forms. Four classes of differential forms are introduced and it is shown that some differential expressions connected in a natural way to quasiregular mappings are members in these classes.

1. Introduction.

Let Ω be a domain in $R^n, n \geq 2$. A mapping $f : \Omega \to R^n$ is called a quasiregular mapping, if $f = (f_1, f_2, \dots, f_n) \in W^1_{n,\mathrm{loc}}(\Omega)$ and if there exists a constant $K \in [1, \infty)$ such that

$$|f'(x)|^n \leq K \det f'(x), \quad \text{for a.e. } x \in \Omega.$$

The following result is well-known in [**Re**] and [**HKM**].

Each of the functions

$$u = f_i(x) \quad (i = 1, 2, \dots, n), \qquad u = \log |f(x)|,$$

is a generalized solution of a quasilinear elliptic equation

$$\operatorname{div} A(x, \nabla u) = 0, \qquad A = (A_1, A_2, \dots, A_n), \tag{1.1}$$

where

$$A_i(x, \xi) = \frac{\partial}{\partial \xi_i} \left(\sum_{i,j=1}^{n} \theta_{i,j}(x) \xi_i \xi_j \right)^{n/2}, \tag{1.2}$$

and $\theta_{i,j}$ are some functions, which can be expressed in terms of the derivative $f'(x)$, and satisfy

$$c_1(K)|\xi|^2 \leq \sum_{i,j}^{n} \theta_{i,j}(x) \xi_i \xi_j \leq c_2(K)|\xi|^2, \tag{1.3}$$

for some constants $c_1(K)$, $c_2(K) > 0$.

This important proposition connects two large sections of analysis namely, quasiregular mapping theory and the theory of partial differential equations. Much progress in quasiregular mapping theory has resulted from the study of Equations (1.1)-(1.3). On the other hand many investigations of solutions of quasilinear equations in the form (1.1)-(1.3) were stimulated by this connection with quasiregular mapping theory.

However, many theorems about quasiregular mappings, obtained in this way for example, in the monograph [**HKM**] do not make use of the special form (1.2) of functions $A_i(x,\xi)$. In fact, what is important is the divergence form of the Equation (1.1) and the existence of constants $c_1(K)$, $c_2(K)$ – the values of these constants are not significant.

We do not know who was the first turning attention to this fact. Possibly, it was first observed in the paper [**Mi**], where the following fact was recorded and used.

Proposition. *The function $u \in W^1_{n,\mathrm{loc}}(\Omega)$ is the solution of some equation of the form* (1.1) *with Condition* (1.3) *if and only if there exists a differential $(n-1)$-form*

$$\theta(x) = \sum_{i=1}^{n} \theta_i(x)\, dx_1 \wedge \ldots \wedge \widehat{dx_i} \wedge \ldots \wedge dx_n \in L^{n/(n-1)}_{\mathrm{loc}}(\Omega),$$

with the properties:

α) *For every function $\phi \in W^1_n(\Omega)$ with compact support we have*

$$\int_\Omega d\phi \wedge \theta = 0\ ,$$

β) *almost everywhere on Ω the following inequalities are true*

$$\nu_1 |du(x)|^n \le *(du(x) \wedge \theta(x))$$

where $*$ denotes the orthogonal complement of a form and

$$|\theta(x)| \le \nu_2 |du(x)|^{n-1},$$

with constants $\nu_1, \nu_2 > 0$.

The proof for this proposition is obvious. The above statement concerning the coordinate functions of a quasiregular mapping f also follows from this proposition. For the case $u = f_1(x)$ we put

$$\theta = df_2 \wedge df_3 \wedge \ldots \wedge df_n\,.$$

In order to show that $u = \log|f(x)|$ satisfies (1.1) it suffices to choose

$$\theta = \frac{1}{|f(x)|^n} \sum_{i=1}^{n} df_1 \wedge \ldots \wedge \widehat{df_i} \wedge \ldots \wedge df_n.$$

Looking carefully at Conditions α) and β) on the function u we see that these conditions are on the 1-form $w = du$ and the $(n-1)$-form θ. Some simple differential forms w, $1 \le \deg w \le n$, satisfying Conditions α) and β)

in domains $\Omega \subset R^n$ were studied in [**Zh1**] and [**Zh2**]. Similar results have been given in [**Iw**], [**FW**], [**MMV1**], [**MMV2**] and [**Sc**].

The purpose of this paper is to study the relations between quasiregular mappings on Riemannian manifolds and differential forms suggested by the aforementioned proposition. We introduce four classes of differential forms and prove membership in these classes of some differential expressions connected in a natural way to quasiregular mappings.

2. Preliminaries.

2.1. Euclidean space. Let X be a topological space. We denote by $\overline{A}$ the closure of a set $A \subset X$, by $\mathrm{int}A$ the interior of A, and by $\partial A = \overline{A} \setminus \mathrm{int}A$ the boundary of A.

By R^n we denote the Euclidean vector space consisting of elements of the form $x = (x^1, \dots, x^n)$, $x^i \in R$, the field of real numbers. In R^n we use the standard inner product $\langle x, y \rangle = \sum_{i=1}^{n} x^i y^i$ and the norm $|x| = \sqrt{\langle x, x \rangle} = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$.

The boundary of the n-dimensional ball with center at x and radius r

$$B(x, r) = \{y \in R^n : |y - x| < r\}$$

is the sphere

$$S(x, r) = \{y \in R^n : |y - x| = r\}.$$

For $E \subset R^n$ and for an integer $k = 1, 2, \dots, n$ we denote by $H_k(E)$ the k-dimensional Hausdorff measure of E.

2.2. Differential forms on R^n. The mutually dual spaces $\bigwedge_k(R^n)$ and $\bigwedge^k(R^n)$ of k-vectors and k-forms (k-covectors) are associated with the Euclidean space R^n. Here one has $\bigwedge^0(R^n) = R = \bigwedge_0(R^n)$, and $\bigwedge_k(R^n) = \{0\} = \bigwedge^k(R^n)$ in the case $k > n$ or $k < 0$. The direct sums

$$\bigwedge_*(R^n) = \oplus_k \bigwedge_k(R^n), \quad \bigwedge^*(R^n) = \oplus_k \bigwedge^k(R^n)$$

generate contravariant and covariant Grassmann algebras on R^n with the exterior multiplication operator $\wedge$.

Let $\omega \in \bigwedge^k(R^n)$ be a covector. We denote by $\Lambda(k, n)$ the set of ordered multi–indices $I = (i_1, i_2, \dots, i_k)$, of integers $1 \leq i_1 < \dots < i_k \leq n$. The form ω can be written in a unique way as the linear combination

$$\omega = \sum_{I \in \Lambda(k,n)} \omega_I \, dx_I.$$

Here ω_I are the coefficients of ω with respect to the standard basis of $\bigwedge^k(R^n)$

$$dx_I = dx_{i_1} \wedge \dots \wedge dx_{i_k}, \quad I = (i_1, i_2, \dots, i_k) \in \Lambda(k, n).$$

Let $I = (i_1, \dots, i_k)$ be a multi–index from $\Lambda(k, n)$. The complement I^* of the multi–index I is the multi-index $I^* = (j_1, \dots, j_{n-k})$ in $\Lambda(n-k, n)$ where the components j_p are in $\{1, \dots, n\} \setminus \{i_1, \dots, i_k\}$. We have

$$dx_I \wedge dx_{I^*} = \sigma dx_1 \wedge \dots \wedge dx_n \tag{2.3}$$

where $\sigma = \sigma(I)$ is the signature of the permutation $(i_1, \dots, i_k, j_1, \dots, j_{n-k})$ in the set $\{1, 2, \dots, n\}$. Note that $\sigma(I^*) = (-1)^{k(n-k)}\sigma(I)$.

Let $dx_I = dx_{i_1} \wedge \dots \wedge dx_{i_k}$ be a differential form of the standard basis of $\bigwedge^k(R^n)$. We set

$$\star dx_I = \sigma(I) dx_{I^*}. \tag{2.4}$$

For $\omega \in \bigwedge^k(R^n)$ with $\omega = \sum_{I \in \Lambda(k,n)} \omega_I \, dx_I$, we set

$$\star\omega = \sum_{I \in \Lambda(k,n)} \omega_I \star dx_I. \tag{2.5}$$

Then $\star\omega$ belongs to $\bigwedge^{n-k}(R^n)$. The differential form $\star\omega$ is called the orthogonal complement of the differential form ω.

The operator $\star : \bigwedge^*(R^n) \to \bigwedge^*(R^n)$, also called Hodge star operator, has the following properties:

If $\alpha, \beta \in \bigwedge^k(R^n)$ and $a, b \in R$, then

$$\star(a\alpha + b\beta) = a \star \alpha + b \star \beta. \tag{2.6}$$

For every ω with $\deg \omega = k$ we have

$$\star(\star\omega) = (-1)^{k(n-k)}\omega. \tag{2.7}$$

We introduce the following notation. Let ω be a differential form of degree k. We set

$$\star^{-1}\omega = (-1)^{k(n-k)} \star \omega. \tag{2.8}$$

The operator $\star^{-1}$ is an inverse to $\star$ in the sense that $\star^{-1}(\star\omega) = \star(\star^{-1}\omega) = \omega$.

The inner or scalar product of the differential forms α and β of the same degree is defined as

$$\langle \alpha, \beta \rangle = \star^{-1}(\alpha \wedge \star\beta) = \star(\alpha \wedge \star\beta). \tag{2.9}$$

The scalar product of differential forms has the usual properties of the scalar product. We set

$$|\omega| = \sqrt{\langle \omega, \omega \rangle}.$$

A differential form ω of degree k is called simple if there are differential forms $\alpha_1, \dots, \alpha_k$ of degree 1 such that

$$\omega = \alpha_1 \wedge \dots \wedge \alpha_k.$$

We note the following useful property of the Euclidean norm: If $\alpha, \beta \in \bigwedge^*(R^n)$, then

$$|\alpha \wedge \beta| \le |\alpha||\beta|,$$

if at least one of the differential forms α, β is simple. If α and β are simple and nonzero, then equality holds if and only if the subspaces associated with α and β are orthogonal. More generally, if $\deg \alpha = p$, $\deg \beta = q$, then

$$|\alpha \wedge \beta| \leq (C^p_{p+q})^{1/2} |\alpha||\beta|, \tag{2.10}$$

see [**Fe**] §1.7.

The linear isomorphism $\operatorname{Hom}\left(\bigwedge_k(R^n), R\right) \simeq \bigwedge^k(R^n)$, that defines the duality of the spaces $\bigwedge_k(R^n)$ and $\bigwedge^k(R^n)$, associates a k-vector with a differential form. A vector $a = (a_1, \dots, a_n) \in R^n$ defines a differential form of degree 1

$$\omega = a_1 dx^1 + a_2 dx^2 + \ldots + a_n dx^n. \tag{2.11}$$

We denote it by Ω_a. Let $u = (u_1, \dots, u_k)$, $u_i \in \bigwedge_1(R^n)$, be a nondegenerated frame. The set of all k-dimensional frames is identified with the set of simple k-vectors. One can prove that the differential form

$$\Omega_u = \Omega_{u_1} \wedge \ldots \wedge \Omega_{u_k}$$

does not depend on the choice of the particular frame from the class of frames equivalent with u. This fact produces a one-to-one correspondence $u \mapsto \Omega_u$ of the set of simple polyvectors onto the set of simple differential forms.

3. Differential forms on Riemannian manifolds.

3.1. Riemannian manifolds. Let $\mathcal{M}$ be an n-dimensional Riemannian manifold with boundary or without boundary. Throughout the sequel we will assume that the manifold $\mathcal{M}$ is orientable and of class C^p where p is at least 3. By $T(\mathcal{M})$ we denote the tangent bundle and by $T_m(\mathcal{M})$ the tangent space at the point $m \in \mathcal{M}$. For each pair of vectors $x, y \in T_m(\mathcal{M})$ the symbol $\langle\, ,\rangle$ denotes their scalar product. The Riemannian connection on $T_m(\mathcal{M})$ gives the natural connection for tensors of every type. This connection preserves the scalar product mentioned above.

Below we shall use standard notation for function classes on manifolds. Thus, for example, the symbol $L^p_{\text{loc}}(D)$ stands for the set of all Lebesgue measurable functions on an open set $D \subset \mathcal{M}$, locally integrable to the power p, $1 \leq p \leq \infty$, on D. The symbol $W^1_{p,\text{loc}}(D)$ stands for the set of functions that have generalized partial derivatives in the sense of Sobolev of class $L^p_{\text{loc}}(D)$ and $\operatorname{Lip}(D)$ denotes the class of all Lipschitz functions on D.

Let $\mathcal{M}$ and $\mathcal{N}$ be Riemannian manifolds of class C^k, $k \geq 3$, and $F : D \to \mathcal{N}$, $D \subset \mathcal{M}$, a mapping. We shall say that $F \in L^p_{\text{loc}}(D)$ if for an arbitrary function $\phi \in C^0(\mathcal{N})$ we have $\phi \circ F \in L^p_{\text{loc}}(D)$. The mapping F is in the class $W^1_{p,\text{loc}}(D)$, if $\phi \circ F \in W^1_{p,\text{loc}}(D)$ for every $\phi \in C^1(\mathcal{N})$.

Let $V(\mathcal{M})$ be a vector bundle on $\mathcal{M}$. Let in the elements of this bundle be given a Euclidean scalar product and let the linear connection on $V(\mathcal{M})$ preserve this scalar product. In this case we may say that $V(\mathcal{M})$ is a Riemannian vector bundle over $\mathcal{M}$.

By $\bigwedge^k(\mathcal{M})$ and $\bigwedge_k(\mathcal{M})$ we denote Riemannian vector bundles $\bigwedge^k(T_m(\mathcal{M}))$ and $\bigwedge_k(T_m(\mathcal{M}))$. The sections of these bundles are the fields of k-covectors (k-forms) and k-vectors, which we shall discuss now in some detail.

3.2. Basic properties of differential forms. Let $x^1, \dots, x^n$ be local coordinates in the neighborhood of a point $m \in \mathcal{M}$. The square of a line element on $\mathcal{M}$ has the following expression in terms of the local coordinates $x^1, \dots, x^n$

$$ds^2 = \sum_{i,j=1}^{n} g_{ij} dx^i \, dx^j .$$

By the symbol g^{ij} we shall denote the contravariant tensor defined by the equality

$$(g^{ik})(g_{kj}) = (\delta^i_j), \quad i, j = 1, \dots, n,$$

where δ^j_i is the Kronecker symbol.

Each section α of the bundle $\bigwedge^k(\mathcal{M})$ (that is a differential form) can be written in terms of the local coordinates $x^1, \dots, x^n$ as the linear combination

$$\alpha = \sum_{I \in \Lambda(k,n)} \alpha_I \, dx_I = \sum_{1 \le i_1 < \ldots < i_k \le n} \alpha_{i_1 \ldots i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k}. \tag{3.3}$$

Let α be a differential form defined on an open set $D \subset \mathcal{M}$. If $\mathcal{F}(D)$ is a class of functions defined on D then we say that the differential form α is in this class provided that $\alpha_I \in \mathcal{F}(D)$. For instance, the differential form α is in the class $L^p(D)$ if all its coefficients are in this class.

The orthogonal complement of a differential form α on a Riemannian manifold $\mathcal{M}$ will be denoted by $\star\alpha$. If $\deg \alpha = 1$ then in the local orthonormal system of coordinates $x^1, \dots, x^n$ at m we can write

$$\star\alpha(m) = \star \sum_{i=1}^{n} \alpha_i(m) \, dx^i = \sum_{i=1}^{n} (-1)^{i-1} \alpha_i(m) \, dx^1 \wedge \ldots \widehat{dx^i} \ldots \wedge dx^n \, ,$$

where the sign $\widehat{\ }$ means that the expression under $\widehat{\ }$ is omitted. We remark that the differential form dv is the volume element on $\mathcal{M}$.

If α, $\deg \alpha = k$, $0 \le k \le n$, is a differential form whose coefficients are in $C^1(\mathcal{M})$ then $d\alpha$, $\deg(d\alpha) = k + 1$, denotes its differential defined by

$$d\alpha = \sum_{I \in \Lambda(k,n)} d\alpha_I \wedge dx_I .$$

The differentiation is a linear operation for which the following properties hold:

If α and β are arbitrary differential form that are differentiable in a domain $U \subset \mathcal{M}$ then

(i) $d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta$,

(ii) $d(d\alpha) = 0$,

where k is the degree of the differential form α.

The operator $\star$ and the exterior differentiation d define the codifferential operator δ by the formula

$$\delta\alpha = (-1)^k \star^{-1} d \star \alpha \tag{3.4}$$

for a differential form α of degree k. Clearly, $\delta\alpha$ is a differential form of degree $k-1$.

Let $\mathcal{M}$ be a compact n-dimensional orientable Riemannian manifold with nonempty piecewise smooth boundary $\partial\mathcal{M}$. The following Stokes formula holds

$$\int\limits_{\partial\mathcal{M}} \alpha = \int\limits_{\mathcal{M}} d\alpha,$$

for an arbitrary form $\alpha \in C^1(\mathcal{M})$, $\deg\alpha = n-1$.

3.5. A differential form α of degree k on the manifold $\mathcal{M}$ with coefficients $\alpha_{i_1 \ldots i_k} \in L^p_{\mathrm{loc}}(\mathcal{M})$ is called weakly closed if for each differential form β, $\deg\beta = k+1$, with

$$\operatorname{supp}\beta \cap \partial\mathcal{M} = \emptyset, \quad \operatorname{supp}\beta = \overline{\{m \in \mathcal{M} : \beta \neq 0\}} \subset \mathcal{M},$$

and with coefficients in the class $W^1_{q,\mathrm{loc}}(\mathcal{M})$, $\frac{1}{p} + \frac{1}{q} = 1$, $1 \leq p, q \leq \infty$, we have

$$\int\limits_{\mathcal{M}} \langle \alpha, \delta\beta \rangle dv = 0. \tag{3.6}$$

For smooth differential forms α Condition (3.6) agrees with the traditional condition of closedness $d\alpha = 0$. In fact, if $\alpha, \beta \in C^1(\mathcal{M}), \operatorname{supp}\beta \cap \partial\mathcal{M} = \emptyset$, then we have

$$\int\limits_{\mathcal{M}} d\alpha \wedge \star\beta = \int\limits_{\mathcal{M}} d(\alpha \wedge \star\beta) + (-1)^{k+1} \int\limits_{\mathcal{M}} \alpha \wedge d \star \beta.$$

Because the differential form β has compact support on the orientable manifold $\mathcal{M}$ the first integral on the right hand side is zero by the Stokes formula. Thus we get

$$\int\limits_{\mathcal{M}} d\alpha \wedge \star\beta = (-1)^{k+1} \int\limits_{\mathcal{M}} \alpha \wedge \star\star^{-1} d \star \beta = \int\limits_{\mathcal{M}} \alpha \wedge \star\delta\beta = \int\limits_{\mathcal{M}} \langle \alpha, \delta\beta \rangle dv.$$

We fix an arbitrary point $m \in \mathcal{M}$ and pass to the local coordinates on $\mathcal{M}$ in a neighborhood of this point. Using Condition (3.6) and the fundamental lemma of the variational calculus, the *du Bois-Reymond Lemma*, we conclude that everywhere in this neighborhood of m the coefficients of the differential form $d\alpha$ are zero. Thus the validity of (3.6) under the given conditions on β is equivalent to the requirement $d\alpha = 0$ understood in the classical sense.

We next introduce the following very useful theorem.

Theorem 3.7. *Let α and β be differential forms, $\beta \in W_q^1(\mathcal{M})$ with a compact support, and $\alpha \in W_{p,\mathrm{loc}}^1(\mathcal{M})$, $1 \le p, q \le \infty$, $\deg\alpha + \deg\beta = n-1$, $1/p + 1/q = 1$. Then*

$$\int\limits_{\mathcal{M}} d\alpha \wedge \beta = (-1)^{\deg\alpha + 1} \int\limits_{\mathcal{M}} \alpha \wedge d\beta. \tag{3.8}$$

In particular, the differential form α is weakly closed if and only if $d\alpha = 0$ a.e. on $\mathcal{M}$.

Proof. Fix α and β with the stated properties. Because the coefficients of the differential form α are in the class $W_{p,\mathrm{loc}}^1(\mathcal{M})$ there exists a sequence $\{\alpha_n\}_{n=1}^{\infty}$ of differential forms with coefficients of class $C^1(\mathcal{M})$ converging in the W_p^1-norm to the coefficients of the differential form α on every compact set $K \subset \mathrm{int}\mathcal{M}$.

Let $\{\beta_n\}_{n=1}^{\infty}$ be a sequence of differential forms of degree $\deg\beta_n = \deg\beta$ in the class $C^1(\mathcal{M})$ having compact supports and converging in the norm of W_q^1 to the form β. We may assume that there exists a smooth submanifold $U \subset\subset \mathcal{M}$ such that $\mathrm{supp}\,\beta_n \subset U$ for all integers n.

The differential forms $\alpha_n \wedge \beta_n$ have compact supports contained in U. The Stokes formula yields

$$\int\limits_{\mathcal{M}} d(\alpha_n \wedge \beta_n) = \int\limits_{U} d(\alpha_n \wedge \beta_n) = 0,$$

and hence

$$\int\limits_{U} d\alpha_n \wedge \beta_n + (-1)^{\deg\alpha} \int\limits_{U} \alpha_n \wedge d\beta_n = 0.$$

We have

$$\int\limits_{U} d\alpha \wedge \beta - \int\limits_{U} d\alpha_n \wedge \beta_n = \int\limits_{U} (d\alpha - d\alpha_n) \wedge \beta + \int\limits_{U} d\alpha_n \wedge (\beta - \beta_n).$$

Therefore, using inequality (2.10) we obtain

$$\begin{aligned}&\left|\int\limits_U d\alpha\wedge\beta-\int\limits_U d\alpha_n\wedge\beta_n\right|\\ &\le\int\limits_U |d(\alpha-\alpha_n)\wedge\beta|dv+\int\limits_U |d\alpha_n\wedge(\beta-\beta_n)|dv\\ &\le C\int\limits_U |d(\alpha-\alpha_n)|\,|\beta|dv+C\int\limits_U |d\alpha_n|\,|\beta-\beta_n|dv\\ &\le C||d(\alpha-\alpha_n)||_{L^p(U)}\,||\beta||_{L^q(U)}+C||d\alpha_n||_{L^p(U)}\,||\beta-\beta_n||_{L^q(U)},\end{aligned}$$

where $C=\max(C_n^{k+1})^{1/2}$ and $k=\deg\alpha$.

Similarly we obtain

$$\begin{aligned}&\left|\int\limits_U \alpha\wedge d\beta-\int\limits_U \alpha_n\wedge d\beta_n\right|\\ &\le C_1||\alpha||_{L^p(U)}\,||d(\beta-\beta_n)||_{L^q(U)}+C_1||\alpha-\alpha_n||_{L^p(U)}\,||d\beta||_{L^q(U)},\end{aligned}$$

where $C_1=(C_n^k)^{1/2}$.

These inequalities easily yield (3.8).

If $d\alpha=0$ a.e. on $\mathcal{M}$ then by (3.8)

$$\int\limits_{\mathcal{M}}\alpha\wedge d\beta=0 \tag{3.9}$$

for an arbitrary differential form $\beta\in W_q^1$ with compact support. This, obviously, implies (3.6).

On the other hand, if we take a weakly closed differential form $\alpha\in W^1_{p,\mathrm{loc}}(\mathcal{M})$ then by (3.8) one has

$$\int\limits_{\mathcal{M}} d\alpha\wedge\beta=0\quad\text{for all }\beta\in W_q^1(\mathcal{M})\quad\text{with}\quad\operatorname{supp}\beta\subset\mathcal{M}.$$

We fix an arbitrary point $m\in\mathcal{M}$, pass again to the local coordinates on $\mathcal{M}$ in a neighborhood of m and use again the *du Bois-Reymond Lemma* to conclude that almost everywhere in this neighborhood the form $d\alpha$ is zero. □

4. The $\mathcal{WT}$-classes of differential forms.

In this section we introduce several classes of differential forms with generalized derivatives which first were presented in [**MMV1**] and [**MMV2**].

These classes are used to study the associated classes of quasilinear elliptic partial differential equations.

Let $\mathcal{M}$ be a Riemannian manifold of class C^3, $\dim \mathcal{M} = n$, with a boundary or without boundary and let

$$w \in L^p_{\mathrm{loc}}(\mathcal{M}),\ \deg w = k,\ 0 \le k \le n,\ p > 1, \tag{4.1}$$

be a weakly closed differential form on $\mathcal{M}$.

Definition 4.2. A differential form w (4.1) is said to be of the class $\mathcal{WT}_1$ on $\mathcal{M}$ if there exists a weakly closed differential form

$$\theta \in L^q_{\mathrm{loc}}(\mathcal{M}),\ \deg \theta = n-k,\ \frac{1}{p} + \frac{1}{q} = 1, \tag{4.3}$$

such that almost everywhere on $\mathcal{M}$ we have

$$\nu_0\ |\theta|^q \le\ \langle w, \star\theta\rangle, \tag{4.4}$$

where ν_0 is a constant.

Definition 4.5. The differential form (4.1) is said to be of the class $\mathcal{WT}_2$ on $\mathcal{M}$ if there exists a differential form (4.3) such that almost everywhere on $\mathcal{M}$ the conditions

$$\nu_1\, |w|^p \le\ \langle w, \star\theta\rangle \tag{4.6}$$

and

$$|\theta| \le \nu_2\, |w|^{p-1} \tag{4.7}$$

are satisfied, with constants $\nu_1, \nu_2 > 0$.

For an arbitrary simple differential form of degree k

$$w = w_1 \wedge \ldots \wedge w_k$$

we set

$$\|w\| = \left(\sum_{i=1}^{k} |w_i|^2\right)^{1/2}.$$

For a simple differential form w we have Hadamard's inequality

$$|w| \le \prod_{i=1}^{k} |w_i|.$$

Taking these and using the inequality between geometric and arithmetic means

$$\left(\prod_{i=1}^{k} |w_i|\right)^{1/k} \le \frac{1}{k}\sum_{i=1}^{k} |w_i| \le \left(\frac{1}{k}\sum_{i=1}^{k} |w_i|^2\right)^{1/2}$$

we obtain

$$|w| \le k^{-\frac{k}{2}}\, \|w\|^k. \tag{4.8}$$

Definition 4.9. A simple differential form of degree k

$$w = w_1 \wedge \ldots \wedge w_k,\ w_i \in L^p_{\mathrm{loc}}(\mathcal{M}),\ 1 \le i \le k,$$

is said to be of the class $\mathcal{WT}_3$ on $\mathcal{M}$ if there is a differential form (4.3) such that almost everywhere on $\mathcal{M}$ the inequality (4.7) holds and

$$\nu_3 \, \|w\|^{kp} \le k^{\frac{kp}{2}} \, \langle w, \star\theta\rangle. \tag{4.10}$$

Definition 4.11. A simple differential form of degree k

$$w = w_1 \wedge \ldots \wedge w_k,\ w_i \in L^p_{\mathrm{loc}}(\mathcal{M}),\ 1 \le i \le k,$$

is said to be of the class $\mathcal{WT}_4$ on $\mathcal{M}$, if there exists a simple differential form (4.3) such that the inequality (4.10) holds almost everywhere on $\mathcal{M}$ and

$$(n-k)^{\frac{-(n-k)}{2}} \|\theta\|^{n-k} \le \nu_4 \, |w|^{p-1}. \tag{4.12}$$

Remark 4.13. Because every differential form of degree 1 is simple, for $k = 1$ the class $\mathcal{WT}_2$ coincides with the class $\mathcal{WT}_3$ while for $k = n-1$ the class $\mathcal{WT}_3$ coincides with $\mathcal{WT}_4$.

Theorem 4.14. *The following inclusions hold between these $\mathcal{WT}$-classes*

$$\mathcal{WT}_4 \subset \mathcal{WT}_3 \subset \mathcal{WT}_2 \subset \mathcal{WT}_1.$$

Proof. The first two relations follow in an obvious way from (4.8). For the proof of the last one it is enough to observe that

$$|\theta|^q = |\theta|^{\frac{p}{p-1}} \le \left(\nu_2^{\frac{1}{p-1}} |w|\right)^p \le \nu_2^{\frac{p}{p-1}} \nu_1^{-1} \, \langle w, \star\theta\rangle.$$

□

Example 4.15. Let v be a differential form of the class $L^2_{\mathrm{loc}}(\mathcal{M})$ with $\deg v = k$, $1 \le k \le n$. Following Hodge [**Ho**] we shall say that the differential form v is harmonic if it is simultaneously weakly closed and weakly coclosed, that is

$$dv = \delta v = 0. \tag{4.16}$$

In particular, if $f \in C^2(\mathcal{M})$ then the differential form df of degree 1 is harmonic if and only if $\Delta f = 0$.

Theorem 4.17. *Let v be a differential form of $L^2_{\mathrm{loc}}(\mathcal{M})$, $\deg v = k$. If v is a harmonic differential form then v is of the class $\mathcal{WT}_2$ on $\mathcal{M}$ with structure constants $p = 2$, $\nu_1 = \nu_2 = 1$.*

Proof. Setting $\theta = \star^{-1} v \in L^2_{\mathrm{loc}}(\mathcal{M})$ we have

$$\langle v, \star\theta\rangle = \langle v, v\rangle = |v|^2$$

and $|\theta| = |v|$. The differential form $\star^{-1} v$ is weakly closed because $\star^{-1} v = (-1)^{k(n-k)} \star v$. Therefore Conditions (4.6) and (4.7) indeed hold with the constants $p = 2$, $\nu_2 = \nu_3 = 1$. □

5. Quasilinear elliptic equations.

Let $\mathcal{M}$ be a Riemannian manifold and let

$$A : \bigwedge^k(T(\mathcal{M})) \to \bigwedge^k(T(\mathcal{M}))$$

be a mapping defined almost everywhere on the k-vector tangent bundle $\bigwedge^k(T(\mathcal{M}))$. We assume that for almost every $m \in \mathcal{M}$ the mapping A is defined on the k-vector tangent space $\bigwedge^k(T_m(\mathcal{M}))$, that is for almost every $m \in \mathcal{M}$ the mapping

$$A(m,\ .\ ) : \xi \in \bigwedge^k(T_m(\mathcal{M})) \to \bigwedge^k(T_m(\mathcal{M}))$$

is defined and continuous. We assume that the mapping $m \mapsto A_m(X)$ is measurable for all measurable k-vector fields X. Suppose that for almost every $m \in \mathcal{M}$ and for all $\xi \in \bigwedge^k(T_m(\mathcal{M}))$ we have

$$\nu_0\, |A(m,\xi)|^p \le \langle \xi, A(m,\xi)\rangle \tag{5.1}$$

with the constants $p > 1$ and $\nu_0 > 0$.

Definition 5.2. A differential form $w \in W^{1,p}_{\mathrm{loc}}(\mathcal{M})$ is said to be A-harmonic if it is a solution of the A-harmonic equation

$$\delta A(m, dw) = 0, \tag{5.3}$$

understood in the weak sense, that is

$$\int\limits_{\mathcal{M}} \langle d\Phi, A(m, dw)\rangle\, dv = 0 \tag{5.4}$$

for all differential forms $\Phi \in W^{1,q}_{\mathrm{loc}}(\mathcal{M})$, $1/p + 1/q = 1$, with $\operatorname{supp}\Phi \cap \partial\mathcal{M} = \emptyset$.

Theorem 5.5. *If the differential form $w \in W^1_{p,\mathrm{loc}}(\mathcal{M})$ is A-harmonic with the property* (5.1) *then the differential form dw is in the class $\mathcal{WT}_1$ on $\mathcal{M}$.*

Proof. Let w, $\deg w = k$ be a solution of (5.3) understood in the weak sense. Let the differential form $\alpha(m)$ be associated with the vector field $A(m, dw)$ at the point m and set $\theta = \star\,\alpha$. The differential form w is weakly closed because of (5.4) and the weak closedness of θ follows from

$$\begin{aligned}(-1)^{nk+1}\int\limits_{\mathcal{M}} \langle \theta, \delta\psi\rangle\, dv &= \int\limits_{\mathcal{M}} \langle \star\,\alpha, \star\, d \star \psi\rangle\, dv \\ &= \int\limits_{\mathcal{M}} \langle \alpha, d \star \psi\rangle\, dv = \int\limits_{\mathcal{M}} \langle A(m, dw), d\phi\rangle\, dv \;=\; 0\end{aligned}$$

for all $\psi = \star^{-1}\phi \in W^{1,q}(\mathcal{M})$ with $\operatorname{supp}\psi \cap \partial\mathcal{M} = \emptyset$. Further, by (5.1) we get

$$\nu_0|\theta|^q = \nu_0|A(m, dw)|^q \le \langle dw, A(m, dw)\rangle = \langle dw, \star\theta\rangle,$$

which guarantees (4.4). □

From now on we assume that the vector field $A(m,\xi)$ satisfies the conditions

$$\nu_1\,|\xi|^p \le \langle \xi, A(m,\xi)\rangle, \tag{5.6}$$

and

$$|A(m,\xi)| \le \nu_2\,|\xi|^{p-1} \tag{5.7}$$

with $p > 1$ and for some constants $\nu_1, \nu_2 > 0$. It is clear that we have $\nu_1 \le \nu_2$.

Theorem 5.8. *A differential form $\omega \in W^{1,p}_{\mathrm{loc}}(\mathcal{M})$ is A-harmonic with properties* (5.6) *and* (5.7) *if and only if $d\omega \in \mathcal{WT}_2$.*

Proof. As is the proof of Theorem 5.5 we define θ. The weak closedness of w and θ follows as above. From (5.6) it follows that

$$\nu_1|dw|^p \le \langle dw, A(m,dw)\rangle = \langle dw, \star\,\theta\rangle$$

and from (5.7)

$$|\theta| = |\star\,\alpha| = |A(m,dw)| \le \nu_2|dw|^{p-1}.$$

Conversely, if $dw \in \mathcal{WT}_2$, then there exists a weakly closed differential form θ (see (4.3)) such that (4.6) and (4.7) are satisfied. With the vector field $a : \mathcal{M} \to \Lambda_k(\mathbb{R})$ associated to the differential form $\alpha = \star\,\theta$ we define

$$A(m,\xi) = \begin{cases} a(m), & \text{for } \xi = dw(m), \\ \xi|\xi|^{p-2}, & \text{for } \xi \ne dw(m). \end{cases} \tag{5.9}$$

The weak closedness of θ ensures that w is a solution of (5.3) understood in the weak sense. Conditions (5.6) and (5.7) for A are satisfied with (4.6) and (4.7). □

6. Quasiregular mappings.

Let $\mathcal{M}$ and $\mathcal{N}$ be Riemannian manifolds of dimension n. A mapping $F : \mathcal{M} \to \mathcal{N}$ of the class $W^1_{n,\mathrm{loc}}(\mathcal{M})$ is called a quasiregular mapping if F satisfies

$$|F'(m)|^n \le K J_F(m) \tag{6.1}$$

almost everywhere on $\mathcal{M}$. Here $F'(m) : T_m(\mathcal{M}) \to T_{F(m)}(\mathcal{N})$ is the formal derivative of $F(m)$, further, $|F'(m)| = \max_{|h|=1} |F'(m)h|$. We denote by $J_F(m)$ the Jacobian of F at the point $m \in \mathcal{M}$, i.e., the determinant of $F'(m)$.

The best constant $K \ge 1$ in the inequality (6.1) is called the outer dilatation of F and denoted by $K_O(F)$. If F is quasiregular then the least constant $K \ge 1$ for which we have

$$J_F(m) \le K l(F'(m))^n$$

almost everywhere on $\mathcal{M}$ is called the inner dilatation of the mapping $F : \mathcal{M} \to \mathcal{N}$ and denoted by $K_I(F)$. Here

$$l(F'(m)) = \min_{|h|=1} |F'(m)h|.$$

The quantity

$$K(F) = \max\{K_O(F), K_I(F)\}$$

is called the maximal dilatation of F and if $K(F) \le K$ then the mapping F is called K-quasiregular.

If $F : \mathcal{M} \to \mathcal{N}$ is a quasiregular homeomorphism then the mapping F is called quasiconformal. In this case the inverse mapping F^{-1} is also quasiconformal in the domain $F(\mathcal{M}) \subset \mathcal{N}$ and $K(F^{-1}) = K(F)$.

Example 6.2. Some basic examples of quasiregular mappings are provided by mappings $F : \mathcal{M} \to \mathcal{N}$ that distort lengths of curves by a bounded factor. Indeed, following [**HKM**], we shall say that a mapping $F : \mathcal{M} \to \mathcal{N}$, $F \in W^1_{1,\text{loc}}(\mathcal{M})$, is an L–BLD mapping if $J_F(m) \ge 0$ almost everywhere on $\mathcal{M}$ and for some constant $L \ge 1$ and for all $h \in T_m(\mathcal{M})$ and almost every $m \in \mathcal{M}$ we have

$$|h|/L \le |F'(m)h| \le L|h|. \tag{6.3}$$

It is readily shown that every L–BLD map is K-quasiregular with $K = L^{2(n-1)}$ ([**HKM**], Lemma 14.80).

Let $\mathcal{A}$ and $\mathcal{B}$ be Riemannian manifolds of dimensions $\dim \mathcal{A} = k$, $\dim \mathcal{B} = n-k$, $1 \le k < n$, and with scalar products $\langle\,,\rangle_A$, $\langle\,,\rangle_B$, respectively. On the Cartesian product $\mathcal{N} = \mathcal{A} \times \mathcal{B}$ we introduce the natural structure of a Riemannian manifold with the scalar product

$$\langle\,,\rangle = \langle\,,\rangle_{\mathcal{A}} + \langle\,,\rangle_{\mathcal{B}}.$$

We denote by $\pi : \mathcal{A} \times \mathcal{B} \to \mathcal{A}$ and $\eta : \mathcal{A} \times \mathcal{B} \to \mathcal{B}$ the natural projections of the manifold $\mathcal{N}$ onto submanifolds.

If $w_{\mathcal{A}}$ and $w_{\mathcal{B}}$ are volume forms on A and $\mathcal{B}$, respectively, then the differential form $w_{\mathcal{N}} = \pi^* w_{\mathcal{A}} \wedge \eta^* w_{\mathcal{B}}$ is a volume form on $\mathcal{N}$.

Theorem 6.4. *Let $F : \mathcal{M} \to \mathcal{N}$ be a quasiregular mapping and let $f = \pi \circ F : \mathcal{M} \to \mathcal{A}$. Then the differential form $f^* w_{\mathcal{A}}$ is of the class $\mathcal{WT}_2$ on $\mathcal{M}$ with the structure constants $p = n/k$, $\nu_1 = \nu_1(n,k,K_O)$ and $\nu_2 = \nu_2(n,k,K_O)$.*

Remark 6.5. From the proof of the theorem it will be clear that the structure constants can be chosen to be

$$\nu_1^{-1} = \left(k + \frac{n-k}{\overline{c}^2}\right)^{-n/2} n^{n/2}\, K_O, \quad \nu_2^{-1} = \underline{c}^{\,n-k},$$

where $\overline{c} = \overline{c}(k, n, K_O)$ and $\underline{c} = \underline{c}(k, n, K_O)$ are, respectively, the greatest and least positive roots of the equation

$$(k\xi^2 + (n-k))^{n/2} - n^{n/2} K_O \, \xi^k = 0. \tag{6.6}$$

Proof. Setting $g = \eta \circ F : \mathcal{M} \to \mathcal{B}$ we choose $\theta = g^* w_{\mathcal{B}}$. The volume form $w_{\mathcal{B}}$ is weakly closed.

In fact, if the mapping g is sufficiently regular then

$$d\theta = dg^* w_{\mathcal{B}} = g^* dw_{\mathcal{B}} = 0.$$

In the general case for the verification of Condition (3.6) we approximate the mapping $g : \mathcal{M} \to \mathcal{B}$ in the norm of W_n^1 by smooth maps g_l, $l = 1, 2, \dots$. Because Condition (3.6) holds for each of the differential forms $g_l^* w_{\mathcal{B}}$, it must hold also for the differential form $g^* w_{\mathcal{B}}$.

The weak closedness of the differential form $f^* w_{\mathcal{A}}$ follows similarly.

Fix a point $m \in \mathcal{M}$, at which the relation (6.1) holds. Set $a = f(m)$, $b = g(m)$. Then

$$T_{F(m)}(\mathcal{N}) = T_a(\mathcal{A}) \times T_b(\mathcal{B}).$$

The computations can be conveniently carried out as follows. We first rewrite Condition (6.1) in the form

$$|F'(m)|^n \le K_O |F^* w_{\mathcal{N}}|, \tag{6.7}$$

where $w_{\mathcal{N}}$ is a volume form on $\mathcal{N}$.

For the points $a \in \mathcal{A}$, $b \in \mathcal{B}$ we choose neighborhoods and local systems of coordinates $y^1, \dots, y^k$, and $y^{k+1}, \dots, y^n$, orthonormal at a and b, respectively. We have

$$\begin{aligned} f^* w_{\mathcal{A}} &= f^*(dy^1 \wedge \ldots \wedge dy^k) = f^* dy^1 \wedge \ldots \wedge f^* dy^k \\ &= df^1 \wedge \ldots \wedge df^k, \quad f^i = y^i \circ f, \quad i = 1, \dots, k. \end{aligned}$$

Because the differential form $w_{\mathcal{A}}$ is simple we obtain by the inequality between the geometric and arithmetic means

$$\begin{aligned} |df^1 \wedge \ldots \wedge df^k|^{1/k} &\le \left(\prod_{i=1}^k |df^i| \right)^{1/k} \\ &\le \frac{1}{k} \sum_{i=1}^k |df^i| \le \left(\frac{1}{k} \sum_{i=1}^k |df^i|^2 \right)^{1/2}. \end{aligned} \tag{6.8}$$

Similarly

$$|dg^{k+1} \wedge \ldots \wedge dg^n|^{1/(n-k)} \le \left(\frac{1}{n-k} \sum_{i=k+1}^n |dg^i|^2 \right)^{1/2}. \tag{6.9}$$

It is not difficult to see that

$$F^* w_{\mathcal{N}} = F^*(\pi^* w_{\mathcal{A}} \wedge \eta^* w_{\mathcal{B}}) = f^* w_{\mathcal{A}} \wedge g^* w_{\mathcal{B}} = f^* w_{\mathcal{A}} \wedge \theta$$

and further that

$$|F^* w_{\mathcal{N}}| = |f^* w_{\mathcal{A}} \wedge g^* w_{\mathcal{B}}| \leq |df^1 \wedge \ldots \wedge df^k||dg^{k+1} \wedge \ldots \wedge dg^n|.$$

We have

$$|dF|^2 = \sum_{i=1}^{k} |df^i|^2 + \sum_{i=k+1}^{n} |dg^i|^2 \ \leq n\, |F'|^2.$$

Therefore we get from (6.7), (6.8) and (6.9)

$$\begin{aligned} &\Big(k|f^* w_{\mathcal{A}}|^{2/k} + (n-k)|g^* w_{\mathcal{B}}|^{2/(n-k)}\Big)^{n/2} \\ &\leq n^{n/2}\, K_O \langle f^* w_{\mathcal{A}}, \star\theta\rangle \leq n^{n/2}\, K_O |f^* w_{\mathcal{A}}||g^* w_{\mathcal{B}}|. \end{aligned}$$

Set

$$\xi = \frac{|f^* w_{\mathcal{A}}|^{1/k}}{|g^* w_{\mathcal{B}}|^{1/(n-k)}}.$$

The preceding relation takes the form

$$\big(k\xi^2 + (n-k)\big)^{n/2} \leq n^{n/2}\, K_O \xi^k.$$

Using the notations $\underline{c}$ and $\overline{c}$ for the least and greatest positive roots of Equation (6.6) we have $\underline{c} \leq \xi \leq \overline{c}$ and

$$\underline{c}|g^* w_{\mathcal{B}}|^{1/(n-k)} \leq |f^* w_{\mathcal{A}}|^{1/k} \leq \overline{c}|g^* w_{\mathcal{B}}|^{1/(n-k)}. \tag{6.10}$$

As above, from (6.10) it follows that

$$|f^* w_{\mathcal{A}}|^{n/k} \leq \Big(k + \frac{n-k}{\overline{c}^2}\Big)^{-n/2} n^{n/2}\, K_O \langle f^* w_{\mathcal{A}}, \star\theta\rangle.$$

Thus Condition (4.6) for the membership of the differential form $f^* w_{\mathcal{A}}$ of degree k in the class $\mathcal{WT}_2$ is indeed satisfied.

To verify Condition (4.7) it is enough to observe that from (6.10) it follows that

$$\underline{c}^{n-k}|\theta| \leq |f^* w_{\mathcal{A}}|^{\frac{n-k}{k}}.$$

□

Let $y^1, y^2, \ldots, y^k$ be an orthonormal system of coordinates in R^k, $1 \leq k \leq n$. Let $\mathcal{A}$ be a domain in R^k and let $\mathcal{B}$ be an $(n-k)$-dimensional Riemannian manifold. We consider the manifold $\mathcal{N} = \mathcal{A} \times \mathcal{B}$.

Let $F = (f^1, f^2, \ldots, f^k, g) : \mathcal{M} \to \mathcal{N}$ be a mapping of the class $W^1_{n,\mathrm{loc}}(\mathcal{M})$ and $g = \eta \circ F$ as defined above. We have $f^* w_{\mathcal{A}} = df^1 \wedge \ldots \wedge df^k$.

Theorem 6.11. *If the mapping F is quasiregular then the differential form $f^* w_{\mathcal{A}}$ is in the class $\mathcal{WT}_3$ on $\mathcal{M}$ with the structure constants $p = n/k$, $\nu_3 = \nu_3(k, n, K_O)$, $\nu_2 = \nu_2(k, n, K_O)$.*

Remark 6.12. We can choose the constants ν_2, ν_3 to be

$$\nu_2 = \underline{c}_1^{k-n}, \quad \nu_3 = \left(1 + \frac{1}{\overline{c}_1^2}\right)^{n/2} n^{-n/2} k^{n/2} K_O^{-1}$$

where $\underline{c}_1$ is the least and $\overline{c}_1$ the greatest positive root of the equation

$$(\xi^2 + 1)^{n/2} - n^{n/2} k^{-k/2} (n-k)^{-(n-k)/2} K_O \, \xi^k = 0. \tag{6.13}$$

Proof. In contrast to the previous case the k-form $f^* w_{\mathcal{A}}$ has now a global coordinate representation. Because the earlier arguments had local character they are applicable to the present case, too. As in the previous case we can choose $\theta = g^* w_{\mathcal{B}}$. Condition (4.7) holds with the same constant. We now proceed to verify Condition (4.10).

Combining (6.7), (6.8) and (6.9) we get

$$\left(\sum_{i=1}^{k} |df^i|^2 + \sum_{i=k+1}^{n} |dg^i|^2\right)^{n/2}$$
$$\leq k^{-k/2}(n-k)^{-(n-k)/2} n^{n/2} K_O \left(\sum_{i=1}^{k} |df^i|^2\right)^{k/2} \left(\sum_{i=k+1}^{n} |dg^i|^2\right)^{(n-k)/2}.$$

Here we set

$$\xi = \left(\frac{\sum_{i=1}^{k} |df^i|^2}{\sum_{i=k+1}^{n} |dg^i|^2}\right)^{1/2}.$$

We then get

$$(\xi^2 + 1)^{n/2} \leq k^{-k/2}(n-k)^{-(n-k)/2} n^{n/2} K_O \xi^k.$$

If $\underline{c}_1, \overline{c}_1$ are, respectively, the least and greatest of the positive roots of (6.13) then

$$\underline{c}_1 \left(\sum_{i=k+1}^{n} |dg^i|^2\right)^{1/2} \leq \left(\sum_{i=1}^{k} |df^i|^2\right)^{1/2} \leq \overline{c}_1 \left(\sum_{i=k+1}^{n} |dg^i|^2\right)^{1/2}. \tag{6.14}$$

From the relations (6.7) and (6.14) it follows that

$$\left(\frac{1}{\overline{c}_1^2} + 1\right)^{n/2} \left(\sum_{i=1}^{k} |df^i|^2\right)^{n/2} \leq n^{n/2} K_O \langle f^* w_{\mathcal{A}}, \star\theta \rangle,$$

which guarantees the truth of (4.10). □

Theorem 6.15. *If the mapping $F : \mathcal{M} \to R^n$ is quasiregular then the differential form $f^* w_{\mathcal{A}} = df^1 \wedge \ldots \wedge df^k$ is of the class $\mathcal{WT}_4$ on $\mathcal{M}$ with the structure constants $p = n/k$, $\nu_3 = \nu_3(k, n, K_O)$, $\nu_4 = \nu_4(k, n, K_O)$.*

Proof. As above we set $\theta = dg^{k+1} \wedge \ldots \wedge dg^n$. Condition (4.10) has already been proved. By (6.7), (6.9) and (6.14) we have

$$(1 + \underline{c}_1^2)^{n/2} \left(\sum_{i=k+1}^{n} |dg^i|^2 \right)^{n/2}$$
$$\leq (n-k)^{-(n-k)/2} n^{n/2} K_O |f^* w_{\mathcal{A}}| \left(\sum_{i=k+1}^{n} |dg^i|^2 \right)^{(n-k)/2} .$$

Therefore

$$\left(\sum_{i=k+1}^{n} |dg^i|^2 \right)^{k/2} \leq (n-k)^{-(n-k)/2} (1 + \underline{c}_1^2)^{-n/2} n^{n/2} K_O |f^* w_{\mathcal{A}}|,$$

which easily yields the desired conclusion. □

Remark 6.16. For the constant ν_3 we can choose the constant of Theorem 6.11 and

$$\nu_4 = \Big((n-k)^{-n/2} (1 + \underline{c}_1^2)^{-n/2} n^{n/2} K_O \Big)^{(n-k)/k} .$$

Theorem 6.17. *Let $f = (f^1, f^2, \ldots, f^{n-1}) : \mathcal{M} \to R^{n-1}$ be a mapping of the class $W^1_{n,\mathrm{loc}}(\mathcal{M})$ and let the fundamental group π_1 of the manifold $\mathcal{M}$ be trivial. The mapping f can be extended to a quasiregular mapping*

$$F = (f, f^n) = (f^1, \ldots, f^{n-1}, f^n) : \mathcal{M} \to R^n$$

if and only if the differential form $w = df^1 \wedge \ldots \wedge df^{n-1}$ of degree $n-1$ is in the class $\mathcal{WT}_4$ on $\mathcal{M}$ with $p = n/(n-1)$.

Proof. We assume that $F = (f, f^n)$ is quasiregular. By Theorem 6.15 the differential form w is in the class $\mathcal{WT}_4$ on $\mathcal{M}$.

Conversely, let w be a differential form of the class $\mathcal{WT}_4$. Then there exists a weakly closed differential form θ, $\deg \theta = 1$, satisfying Conditions (4.10) and (4.12). Because $\pi_1 = \{e\}$ there exists an injective function $f^n : \mathcal{M} \to R^1$ such that $df^n = \theta$. From (4.10) we get

$$\nu_3 \left(\sum_{i=1}^{n-1} |df^i|^2 \right)^{n/2} \leq (n-1)^{n/2} |df^1 \wedge \ldots \wedge df^n|.$$

Condition (4.12) implies

$$\nu_4^{-1}|df^n| \le |df^1 \wedge \dots df^{n-1}|^{1/(n-1)} \le \left(\frac{1}{n-1}\sum_{i=1}^{n-1}|df^i|^2\right)^{1/2}.$$

Thus we get

$$\begin{aligned}\left(\sum_{i=1}^{n}|df^i|^2\right)^{n/2} &\le \left(\sum_{i=1}^{n-1}|df^i|^2 + \frac{\nu_4^2}{n-1}\sum_{i=1}^{n-1}|df^i|^2\right)^{n/2} \\ &\le \left(1+\frac{\nu_4^2}{n-1}\right)^{n/2}\frac{1}{\nu_3}(n-1)^{n/2}|df^1 \wedge \dots \wedge df^n|,\end{aligned}$$

which implies (6.1) with the constant

$$K_O = (n-1+\nu_4^2)^{n/2} n^{-n/2}\nu_3^{-1}.$$

□

References

[Fe] H. Federer, *Geometric Measure Theory*, Die Grundlehren der math. Wiss., Vol. 153, Springer-Verlag, Berlin-Heidelberg-New York, 1969, MR 41 #1976, Zbl 176.00801.

[FW] D. Franke and R. Wisk, *Quasireguläre abbildungen und p-harmonische tensoren*, preprint A 1/95, Freie Universität Berlin, 1995.

[HKM] J. Heinonen, T. Kilpeläinen and O. Martio, *Nonlinear Potential Theory of Degenerate Elliptic Equations*, Clarendon Press, 1993, MR 94e:31003, Zbl 780.31001.

[Ho] W.V.D. Hodge, *The Theory and Applications of Harmonic Integrals*, 2nd ed., Cambridge Univ. Press, 1952, MR 14,500b, Zbl 048.15702.

[Iw] T. Iwaniec, *p-Harmonic tensors and quasiregular mappings*, Ann. of Math., **136** (1992), 589-624, MR 94d:30034, Zbl 785.30009.

[MMV1] O. Martio, V.M. Miklyukov and M. Vuorinen, *Phragmén-Lindelöf's principle, differential forms, and quasiregular mappings on Riemannian manifolds*, Manuscript, 1995, 1-127.

[MMV2] ———, *Differential forms and quasiregular mappings on Riemannian manifolds*, 16th Rolf Nevanlinna Colloquium, Eds.: Laine/Martio, Walter de Gruyter & Co., Berlin-New York, (1995), 151-159, MR 98a:58146, Zbl 899.58065.

[Mi] V.M. Miklyukov, *Asymptotic properties of subsolutions of quasilinear equations of elliptic type and mappings with bounded distortion*, Mat. Sb., **11** (1980), 42-66; Engl. transl. in Math. USSR Sb., **39** (1981), 37-60, MR 81j:35037, Zbl 462.35039.

[Re] Yu.G. Reshetnyak, *Space Mappings with Bounded Distortion*, Translations of Mathematical Monographs, 73, American Mathematical Society, Providence, RI, 1989, MR 90d:30067, Zbl 667.30018.

[Sc] C. Scott, *L^p theory of differential forms on manifolds*, Trans. Amer. Math. Soc., **374** (1995), 2075-2096, MR 95i:58009, Zbl 849.58002.

[Zh1] Yu.A. Zhuravleva, *Qualitative properties of solutions of differential inequalities of a special form*, Sibirsk. Mat. Zh., **33**(**3**) (1992), 78-90, MR 93h:35082, Zbl 808.35177.

[Zh2] ______, *Phragmén-Lindelöf principle for differential inequalities of special type*, Dissertation, Volgograd, (1993), 1-115 (Russian).

Received February 15, 2000.

Fachbereich Mathematik und Informatik
Freie Universität Berlin
Arnimallee 2-6,
14195 Berlin
Germany

Mathematics Department
P.O. Box 4 (Yliopistonkatu 5)
FIN-00014 University of Helsinki
Finland

ul. Tulaka 2 apt. 120
Volgograd 400119
Russia

Mathematics Department
P.O. Box 4 (Yliopistonkatu 5)
FIN-00014 University of Helsinki
Finland
E-mail address: vuorinen@csc.fi

Fachbereich Mathematik und Informatik
Freie Universität Berlin
Arnimallee 2-6,
14195 Berlin
Germany

PACIFIC JOURNAL OF MATHEMATICS
Vol. 202, No. 1, 2002

EGGERT'S CONJECTURE ON THE DIMENSIONS OF NILPOTENT ALGEBRAS

Lakhdar Hammoudi

In this paper we prove that for a finite dimensional commutative nilpotent algebra A over a field of prime characteristic $p > 0$, $\dim A \geq p \dim A^{(p)}$, where $A^{(p)}$ is the subalgebra of A generated by the elements x^p. In particular, this solves Eggert's conjecture.

1. Introduction.

In 1971, Eggert [**2**] conjectured that for a finite commutative nilpotent algebra A over a field $\mathbb{K}$ of prime characteristic $p > 0$, $\dim A \geq p \dim A^{(p)}$, where $A^{(p)}$ is the subalgebra of A generated by all the elements x^p, $x \in A$ and $\dim A$, $\dim A^{(p)}$ denote the dimensions of A and $A^{(p)}$ as vector spaces over $\mathbb{K}$.

In [**3**], Stack conjectures that $\dim A \geq p \dim A^{(p)}$ is true for every finite dimensional nilpotent algebra A over $\mathbb{K}$. We point out that some particular cases of Eggert's conjecture have been proved in [**1**, **2**, **3**, **4**].

Here we prove the conjecture for finite dimensional commutative nilpotent algebras. This combined with the results of [**2**] completely describe the group of units of A and the problem set in [**1**]: "When a finite abelian group is isomorphic to the group of units of some finite commutative nilpotent algebras?" is solved. Recall that the group of units of A is the set A with the following operation: $x \cdot y = x + y + xy$, $\forall x, y \in A$.

The author would like to thank the referee and Prof. B. Magurn along with the Algebra Seminar team at Miami university for valuable suggestions and comments.

2. Results.

Our main result is the following:

Theorem. *Let A be a finite dimensional commutative nilpotent algebra over a field $\mathbb{K}$ of characteristic $p > 0$ and let $A^{(p)}$ be the subalgebra of A generated by all the elements x^p, $x \in A$. Then* $\dim A \geq p \dim A^{(p)}$.

To prove the theorem we need an easy lemma on the partition of some sets in $\mathbb{Z}^d_{\geq 0}$ of d-tuples ($d > 0$) of nonnegative integers. Let $\alpha = (\alpha_1, \dots, \alpha_d)$

and $\beta = (\beta_1, \dots, \beta_d)$ be in $\mathbb{Z}^d_{\geq 0}$. Define $\alpha > \beta$ if in the difference $\alpha - \beta = (\alpha_1 - \beta_1, \dots, \alpha_d - \beta_d)$, the left-most nonzero entry is positive and all other entries to the right are nonnegative. It is easy to prove that $>$ is in fact a partial order on $\mathbb{Z}^d_{\geq 0}$, which is compatible with the addition.

Lemma 1. *Let* $(n_1, n_2, \dots, n_d) = n \in \mathbb{Z}^d_{\geq 0}$ *be a fixed d-tuple such that* $(0, \dots, 0, 0) \neq n$ *and consider the following subsets of* $\mathbb{Z}^d_{\geq 0}$*:*

$$\mathbb{Z}^d_{\geq 0}(n) = \{\alpha,\ (0, \dots, 0, 0) \neq \alpha \leq n\},$$

$\mathbb{Z}^d_{\geq 0}(i_1, \dots, i_{d-1}) = \{(i_1, i_2, \dots, i_{d-1}, j),\ 1 \leq j \leq n_d\}, \quad 0 \leq i_k \leq n_k,\ 1 \leq k \leq d-1,$

$$\mathbb{Z}^d_{\geq 0}(0) = \{(i_1, i_2, \dots, i_{d-1}, 0),\ (i_1, i_2, \dots, i_{d-1}, 0) \in \mathbb{Z}^d_{\geq 0}(n)\}.$$

Then the sets $\mathbb{Z}^d_{\geq 0}(i_1, \dots, i_{d-1})$*, and* $\mathbb{Z}^d_{\geq 0}(0)$ *form a partition of* $\mathbb{Z}^d_{\geq 0}(n)$*.*

The proof of the theorem requires also the following lemma due to Bautista [**1**, Proposition 2.1, p. 15]. For completness, we will give a sketch of a proof of this result.

Lemma 2. *Let A be a commutative nilpotent algebra over a field* $\mathbb{K}$ *generated by* $X_1, \dots, X_d$*. Let* $(\alpha_1, \dots, \alpha_d)$ *be an element of* $\mathbb{Z}^d_{\geq 0}$ *such that* $X_1^{\alpha_1} \cdots X_d^{\alpha_d} \neq 0$ *but* $\forall (\beta_1, \dots, \beta_d) \in \mathbb{Z}^d_{\geq 0}, (\beta_1, \dots, \beta_d) > (\alpha_1, \dots, \alpha_d)$*,* $X_1^{\beta_1} \cdots X_d^{\beta_d} = 0$*. Then for the set of ordered d-tuples*

$$S = \left\{(i_1, \dots, i_d) \in \mathbb{Z}^d_{\geq 0};\ (\alpha_1, \dots, \alpha_d) - (i_1, \dots, i_d) \in \mathbb{Z}^d_{\geq 0}\right\},$$

$\{X_1^{i_1} \cdots X_d^{i_d};\ (i_1, \dots, i_d) \in S\}$ *is linearly independent.*

Sketch of Proof. Suppose that the family

$$\left\{X_1^{i_1} \cdots X_d^{i_d};\ (i_1, \dots, i_d) \in \mathbb{Z}^d_{\geq 0};\ (\alpha_1, \dots, \alpha_d) - (i_1, \dots, i_d) \in \mathbb{Z}^d_{\geq 0}\right\}$$

is linearly dependent. Then there exists a set of nonzero elements $\lambda_{i_1, \dots, i_d} \in \mathbb{K}$ such that $\sum_{\alpha - I \in \mathbb{Z}^d_{\geq 0}} \lambda_{i_1, \dots, i_d} X_1^{i_1} \cdots X_d^{i_d} = 0$, $\alpha = (\alpha_1, \dots, \alpha_d)$, $I = (i_1, \dots, i_d)$.

Let $L = (l_1, \dots, l_d)$ be a minimal element such that $\lambda_{l_1, \dots, l_d} \neq 0$. Then

$$\lambda_{l_1, \dots, l_d} X_1^{l_1} \cdots X_d^{l_d} + \sum_{I > L} \lambda_{i_1, \dots, i_d} X_1^{i_1} \cdots X_d^{i_d} = 0.$$

By multiplying on the right by $X_1^{(\alpha_1 - l_1)} \cdots X_d^{(\alpha_d - l_d)}$ and using the commutativity of A, we obtain:

$$\lambda_{l_1, \dots, l_d} X_1^{\alpha_1} \cdots X_d^{\alpha_d} + \sum_{I > L} \lambda_{i_1, \dots, i_d} X_1^{i_1 + (\alpha_1 - l_1)} \cdots X_d^{i_d + (\alpha_d - l_d)} = 0.$$

However, it is easy to see that $(i_1 + \alpha_1 - l_1, \dots, i_d + \alpha_d - l_d) > (\alpha_1, \dots, \alpha_d)$.

Thus,

$$\sum_{I>L} \lambda_{i_1,\dots,i_d} X_1^{i_1+(\alpha_1-l_1)} \cdots X_d^{i_d+(\alpha_d-l_d)} = 0.$$

So, $\lambda_{l_1,\dots,l_d} X_1^{\alpha_1} \cdots X_d^{\alpha_d} = 0$. But, $\lambda_{l_1,\dots,l_d} \neq 0$. Thus, $X_1^{\alpha_1} \cdots X_d^{\alpha_d} = 0$. This contradicts our hypothesis and proves the lemma.

Lemma 3. *Let A be a commutative nilpotent algebra over a field $\mathbb{K}$ generated by d elements $X_1, \dots, X_d$. Suppose that A cannot be generated by $d-1$ elements. Let $\mathcal{B} = \{X_1^{i_1} \cdots X_d^{i_d}, (i_1, i_2, \dots, i_d) \in \mathbb{Z}^d_{\geq 0}$, with the convention $X_k^0 = 1$, $1 \leq k \leq d\}$ be a basis of A as a vector space over $\mathbb{K}$. Then $X_d \in \mathcal{B}$ and some of the basis $\mathcal{B}$ are such that, if for some $(j_1, \dots, j_d)$, $j_d \geq 2$, $X_1^{j_1} \cdots X_d^{j_d} \in \mathcal{B}$ then $X_1^{j_1} \cdots X_{d-1}^{j_{d-1}} X_d^{j_d-1} \in \mathcal{B}$.*

Proof. Suppose that $X_d \notin \mathcal{B}$ and let us write it as a linear combination of elements of $\mathcal{B}$, $X_d = \sum_{i_1,\dots,i_d} \lambda_{i_1,\dots,i_d} X_1^{i_1} \cdots X_d^{i_d}$, $\lambda_{i_1,\dots,i_d} \in \mathbb{K}$. Since A is not generated by $d-1$ elements, for some i_d we have $i_d \geq 1$. So, one can write

$$X_d = \left(\sum_{i_1,\dots,i_d} \lambda_{i_1,\dots,i_d} X_1^{i_1} \cdots X_d^{i_d-1} \right) \left(\sum_{i_1,\dots,i_d} \lambda_{i_1,\dots,i_d} X_1^{i_1} \cdots X_d^{i_d} \right).$$

Since A is commutative and nilpotent, by repeating the above process we can write X_d as a linear combination of monomials in $X_1, \dots, X_{d-1}$. Thus A is generated by $d-1$ elements. This contradiction proves our assertion, $X_d \in \mathcal{B}$.

We prove now our second assertion. It is easy to see that $X_1^{j_1} \cdots X_d^{j_d} \in \mathcal{B}$ implies that there exists $(\alpha_1, \dots, \alpha_d) \in \mathbb{Z}^d_{\geq 0}$ satisfying the hypothesis of Lemma 2 such that

$$(\alpha_1, \dots, \alpha_d) > (j_1, \dots, j_d) \text{ and } (\alpha_1 - j_1, \dots, \alpha_d - j_d) \in \mathbb{Z}^d_{\geq 0}.$$

But $(j_1, \dots, j_d) > (j_1, \dots, j_{d-1}, j_d - 1)$. So, $(\alpha_1 - j_1, \dots, \alpha_{d-1} - j_{d-1}, \alpha_d - j_d - 1) \in \mathbb{Z}^d_{\geq 0}$. Thus, Lemma 2 applies here.

Suppose now that $X_1^{j_1} \cdots X_{d-1}^{j_{d-1}} X_d^{j_d-1} \notin \mathcal{B}$. Then $\{X_1^{j_1} \cdots X_{d-1}^{j_{d-1}} X_d^{j_d-1}, \mathcal{B}\}$ is linearly dependent which contradicts the preceeding lemma.

Proof of the Theorem. We prove our theorem by induction on the number l of generators of the algebra A.

We first prove the conjecture for $l = 1$. Let X be a generator of A and $m+1$ be the degree of nilpotency of X. Then $\{X, X^2, \dots, X^m\}$ is a basis for the vector space A and since A is commutative over a field of characteristic p, $\{X^p, \dots, X^{pk}\}$ is a basis of $A^{(p)}$. But the fact that $m+1$ is the degree of nilpotency of X yields to $m \geq pk$. So, $\dim A = m \geq pk = p \dim A^{(p)}$.

Suppose that the theorem is proved for every algebra generated by l elements, $l \leq d-1$ and consider a finite dimensional commutative nilpotent algebra A over $\mathbb{K}$ generated by d elements, $X_1, \dots, X_d$. Since A is nilpotent, there exists a d-tuple $(n_1, n_2, \dots, n_d) = n \in \mathbb{Z}^d_{\geq 0}$ such that $n_1+1, \dots, n_d+1$ are the degrees of nilpotency of $X_1, \dots, X_d$ respectively. Since A is commutative over a field of characteristic p, as vector spaces over $\mathbb{K}$, A and $A^{(p)}$ are generated by the monomials of the form $\{X_1^{\beta_1} \cdots X_d^{\beta_d}, (\beta_1, \dots, \beta_d) \in \mathbb{Z}^d_{\geq 0}$, where $X_i^0 = 1\}$ and $X_1^{p\beta_1} \cdots X_d^{p\beta_d}$ respectively. So, one can extract a basis $\mathcal{B}$ of $A^{(p)}$ from the last cited monomials. Let $\bar{\mathcal{B}}$ be a basis of A obtained by completing $\mathcal{B}$. Let $\mathbb{Z}^d_{\geq 0}(\bar{\mathcal{B}})$ be the set of all d-tuples $(\alpha_1, \dots, \alpha_d) \in \mathbb{Z}^d_{\geq 0}$ such that $X_1^{\alpha_1} \cdots X_d^{\alpha_d} \in \bar{\mathcal{B}}$ and denote by $\mathbb{Z}^d_{\geq 0}(\mathcal{B})$ the set of all d-tuples $(\alpha_1, \dots, \alpha_d) \in \mathbb{Z}^d_{\geq 0}$ such that $X_1^{\alpha_1} \cdots X_d^{\alpha_d} \in \mathcal{B}$.

With these notations, $\dim A \geq p \dim A^{(p)}$ is the same as $\#\mathbb{Z}^d_{\geq 0}(\bar{\mathcal{B}}) \geq p\#\mathbb{Z}^d_{\geq 0}(\mathcal{B})$, where $\#Y$ is the number of the elements of the set Y.

Let R be the subalgebra of A generated by $\{X_1, \dots, X_{d-1}\}$. Then by the hypothesis of induction, $\dim R \geq p \dim R^{(p)}$. But, $\dim R = \#(\mathbb{Z}^d_{\geq 0}(\bar{\mathcal{B}}) \cap \mathbb{Z}^d_{\geq 0}(0))$ and $\dim R^{(p)} = \#(\mathbb{Z}^d_{\geq 0}(\mathcal{B}) \cap \mathbb{Z}^d_{\geq 0}(0))$. On the other hand, since $\mathbb{Z}^d_{\geq 0}(\bar{\mathcal{B}})$ and $\mathbb{Z}^d_{\geq 0}(\mathcal{B})$ are included in $\mathbb{Z}^d_{\geq 0}(n)$, by Lemma 1 we have:

$$\mathbb{Z}^d_{\geq 0}(\bar{\mathcal{B}}) = \left(\bigcup_{i_1, \dots, i_{d-1}} \mathbb{Z}^d_{\geq 0}(\bar{\mathcal{B}}) \cap \mathbb{Z}^d_{\geq 0}(i_1, \dots, i_{d-1}) \right) \bigcup \left(\mathbb{Z}^d_{\geq 0}(\bar{\mathcal{B}}) \cap \mathbb{Z}^d_{\geq 0}(0) \right)$$

$$\mathbb{Z}^d_{\geq 0}(\mathcal{B}) = \left(\bigcup_{i_1, \dots, i_{d-1}} \mathbb{Z}^d_{\geq 0}(\mathcal{B}) \cap \mathbb{Z}^d_{\geq 0}(i_1, \dots, i_{d-1}) \right) \bigcup \left(\mathbb{Z}^d_{\geq 0}(\mathcal{B}) \cap \mathbb{Z}^d_{\geq 0}(0) \right).$$

Also, by Lemma 1 we have partitions of $\mathbb{Z}^d_{\geq 0}(\bar{\mathcal{B}})$ and $\mathbb{Z}^d_{\geq 0}(\mathcal{B})$. Thus, we only need to prove that

$$\# \bigcup_{i_1, \dots, i_{d-1}} \left(\mathbb{Z}^d_{\geq 0}(\bar{\mathcal{B}}) \cap \mathbb{Z}^d_{\geq 0}(i_1, \dots, i_{d-1}) \right) \geq p\# \bigcup_{i_1, \dots, i_{d-1}} \left(\mathbb{Z}^d_{\geq 0}(\mathcal{B}) \cap \mathbb{Z}^d_{\geq 0}(i_1, \dots, i_{d-1}) \right).$$

Moreover, since we have a disjoint union of sets, we prove that

$$\# \left(\mathbb{Z}^d_{\geq 0}(\bar{\mathcal{B}}) \cap \mathbb{Z}^d_{\geq 0}(i_1, \dots, i_{d-1}) \right) \geq p\# \left(\mathbb{Z}^d_{\geq 0}(\mathcal{B}) \cap \mathbb{Z}^d_{\geq 0}(i_1, \dots, i_{d-1}) \right).$$

Fix $(i_1, \dots, i_{d-1})$ and let j be the greatest integer such that: $X_1^{i_1} \cdots X_{d-1}^{i_{d-1}} X_d^j \in \bar{\mathcal{B}}$ (i.e., $(i_1, \dots, i_{d-1}, j) \in \mathbb{Z}^d_{\geq 0}(\bar{\mathcal{B}})$).

If $j = 0$ or $j = 1$ then $\mathbb{Z}^d_{\geq 0}(\mathcal{B}) \cap \mathbb{Z}^d_{\geq 0}(i_1, \dots, i_{d-1}) = \emptyset$ and our claim is obvious.

If $j \geq 2$ then by Lemma 3, $(i_1, \ldots, i_{d-1}, k) \in \mathbb{Z}^d_{\geq 0}(\bar{\mathcal{B}})$, $\forall k$, $1 \leq k \leq j$ and so, by the choice of the integer j,

$$\# \left(\mathbb{Z}^d_{\geq 0}(\bar{\mathcal{B}}) \cap \mathbb{Z}^d_{\geq 0}(i_1, \ldots, i_{d-1}) \right) = j.$$

On the other hand

$$\mathbb{Z}^d_{\geq 0}(\mathcal{B}) \cap \mathbb{Z}^d_{\geq 0}(i_1, \ldots, i_{d-1}) = \begin{cases} \emptyset \\ \text{or} \\ \{(i_1, \ldots, i_{d-1}, pk),\ 1 \leq pk \leq j\}. \end{cases}$$

The first case is obvious and in the second as for an algebra generated by one element, we have

$$p\# \left(\mathbb{Z}^d_{\geq 0}(\mathcal{B}) \cap \mathbb{Z}^d_{\geq 0}(i_1, \ldots, i_{d-1}) \right) = pt \leq j.$$

This ends the proof of the theorem.

References

[1] R. Bautista, *Units of finite algebras*, An. Inst. Mat. Univ. Nac. Autónoma, México, **16**(**2**) (1976), 1-78 (in Spanish), MR 58 #11011, Zbl 402.16029.

[2] N. Eggert, *Quasi regular groups of finite commutative nilpotent algebras*, Pacific J. Math., **36** (1971), 631-634, MR 44 #262, Zbl 213.32004.

[3] C. Stack, *Dimensions of nilpotent algebras over fields of prime characteristic*, Pacific J. Math., **176** (1996), 263-266, MR 97m:16037, Zbl 876.16013.

[4] ______, *Some results on the structure of finite nilpotent algebras over fields of prime characteristic*, J. Combin. Math. Combin. Comput., **28** (1998), 327-335, MR 2000a:13031, Zbl 922.16012.

Received July 30, 1998 and revised April 19, 2000.

Department of Mathematics and Statistics Miami University
Oxford, Ohio 45056
E-mail address: hammoul@muohio.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 202, No. 1, 2002

SIGNATURES OF LEFSCHETZ FIBRATIONS

Burak Ozbagci

Let M be a smooth 4-manifold which admits a Lefschetz fibration over D^2 or S^2. We develop an algorithm to compute the signature of M using the global monodromy of this fibration. As a corollary we prove that there is no hyperelliptic Lefschetz fibration over S^2 with only reducible singular fibers.

0. Introduction.

The signature of a smooth 4-manifold which admits a hyperelliptic Lefschetz fibration of genus g over a closed surface can be computed using the *local signature formula* given by Matsumoto ([**M1**], [**M2**]) for $g = 1, 2$ and more recently extended by Endo [**E**] for $g \geq 3$.

In this paper we present an algorithm to compute the signature of a smooth 4-manifold which admits an arbitrary (not necessarily hyperelliptic) Lefschetz fibration of any genus over D^2 or S^2. A Lefschetz fibration on a smooth 4-manifold M gives rise to a handlebody description of M, which is determined by a sequence of vanishing cycles. We use this handlebody description [**K**] and Wall's nonadditivity formula for signatures [**W**] to compute the signature of M. Hence we calculate a *'signature contribution'* corresponding to each singular fiber of the given fibration on M.

As a corollary we prove that "*there is no hyperelliptic Lefschetz fibration over S^2 with only reducible singular fibers.*" After we proved and announced this result the general case (not assuming the hyperellipticity) was proved independently by Li [**L**], Smith [**Sm**] and Stipsicz [**St3**] all using a result of this paper (cf. Corollary 7).

Recent results in symplectic topology show that Lefschetz fibrations provide a topological characterization of symplectic 4-manifolds: Donaldson [**D**] has shown that, after perhaps blowing up, a closed symplectic 4-manifold admits a Lefschetz fibration over S^2, and conversely Gompf [**GS**] has shown that most Lefschetz fibrations are symplectic — the exceptions all have fiber-genus one and are blow-ups of torus fibrations with no critical points. Hence by computing the signatures of Lefschetz fibrations we hope to attack some of the problems in the *geography* of symplectic 4-manifolds ([**St1**], [**St2**], [**GS**]).

We also prove that the signature of a smooth 4-manifold which admits a hyperelliptic Lefschetz fibration of genus $g \leq 3$ over S^2 is nonpositive. It is conjectured that this is true for all genus g Lefschetz fibrations over S^2.

We want to point out that despite the fact that the vanishing cycles are defined up to isotopy, our technique shows that the signature of a 4-manifold which admits a Lefschetz fibration depends only on the algebraic data given by the homology classes of the vanishing cycles.

In [**Sm**], Smith gave an elegant signature formula using the geometry of Lefschetz fibrations. Even though his formula is in a closed form, it seems impossible to actually compute the signature using his formula.

Acknowledgement. The author would like to thank John Etnyre, Terry Fuller, Ludmil Katzarkov, Ron Stern and Richard Wentworth for helpful conversations. In particular, the author would like to thank Terry Fuller for stimulating his interest in Lefschetz fibrations. The author would also like to thank Yukio Matsumoto for informing him about the preprint [**E**] by Hisaaki Endo. The author would like to thank Andras Stipsicz for commenting on earlier versions of this paper. Finally, the author wishes to express his gratitute to Ron Stern for encouragement and many useful discussions.

1. Preliminaries.

1.1. Mapping class groups. Let Σ_g be a closed oriented surface of genus g. Let $\mathrm{Diff}^+(\Sigma_g)$ be the group of all orientation preserving self diffeomorphisms of Σ_g. Let $\mathrm{Diff}_0^+(\Sigma_g)$ be the subgroup of $\mathrm{Diff}^+(\Sigma_g)$ consisting of all self diffeomorphisms isotopic to the identity. Then we define the *mapping class group* of genus g as

$$\mathcal{M}_g = \mathrm{Diff}^+(\Sigma_g)/\mathrm{Diff}_0^+(\Sigma_g).$$

The *hyperelliptic mapping class group* $\mathcal{H}_g$ of genus g is defined as the subgroup of $\mathcal{M}_g$ which consists of all isotopy classes commuting with the isotopy class of the hyperelliptic involution $\iota : \Sigma_g \longrightarrow \Sigma_g$.

It is known that the hyperelliptic mapping class group $\mathcal{H}_g$ agrees with the mapping class group $\mathcal{M}_g$ for $g = 1, 2$ (cf. [**BH**]).

A positive (or right-handed) *Dehn twist* $D(\alpha) : \Sigma_g \to \Sigma_g$ about a simple closed curve α is a diffeomorphism obtained by cutting Σ_g along α, twisting $360°$ to the right and regluing. Note that the positive Dehn twist $D(\alpha)$ is determined up to isotopy by α and is independent of the orientation on α.

It is well-known that the mapping class group $\mathcal{M}_g$ is generated by Dehn twists.

We will use the functional notation for the products in $\mathcal{M}_g$, e.g., $D(\beta)D(\alpha)$ will denote the composition where we apply $D(\alpha)$ first and then $D(\beta)$.

1.2. Smooth Lefschetz fibrations.

Let M be a compact, oriented smooth 4-manifold, and let B be a compact, oriented 2-manifold. A proper smooth map $f : M \to B$ is a smooth Lefschetz fibration if there exist points $b_1, \ldots , b_m \in \text{interior}(B)$ such that

(1) $\{b_1, \ldots , b_m\}$ are the critical values of f, with $p_i \in f^{-1}(b_i)$ a unique critical point of f, for each i, and
(2) about each b_i and p_i, there are local complex coordinate charts agreeing with the orientations of M and B such that locally f can be expressed as $f(z_1, z_2) = z_1^2 + z_2^2$.

Remark. An *achiral* Lefschetz fibration is a fibration which satisfies (1) and (2) above without requiring the coincidence of the canonical orientation determined by (z_1, z_2) and the orientation of M.

It is a consequence of the definition of a smooth Lefschetz fibration that

$$f|_{f^{-1}(B-\{b_1,\ldots,b_m\})} : f^{-1}(B - \{b_1, \ldots , b_m\}) \to B - \{b_1, \ldots , b_m\}$$

is a smooth fiber bundle over $B - \{b_1, \ldots , b_m\}$ with fiber diffeomorphic to a 2-manifold Σ_g, and so we refer to f (and sometimes also the manifold M) as a *genus g Lefschetz fibration* (or a *Lefschetz fibration of genus g*). Two genus g Lefschetz fibrations $f : M \to B$ and $f' : M' \to B'$ are *equivalent* if there are diffeomorphisms $\Phi : M \to M'$ and $\phi : B \to B'$ such that $f'\Phi = \phi f$.

We always assume that our Lefschetz fibrations are *relatively minimal*, namely that no fiber contains an embedded 2-sphere of self-intersection number -1. We also assume that there is at least one singular fiber in each fibration.

If $f : M \to D^2$ is a smooth genus g Lefschetz fibration, then we can use this fibration to produce a handlebody description of M. We select a regular value $b_0 \in \text{interior}(D^2)$ of f, an identification $f^{-1}(b_0) \cong \Sigma_g$, and a collection of arcs s_i in interior(D^2) with each s_i connecting b_0 to b_i, and otherwise disjoint from the other arcs. We also assume that the critical values are indexed so that the arcs $s_1, \ldots , s_m$ appear in order as we travel counter-clockwise in a small circle about b_0. Let $V_0, \ldots , V_m$ denote a collection of small disjoint open disks with $b_i \in V_i$ for each i.

To build our description of M, we observe first that $f^{-1}(V_0) \cong \Sigma_g \times D^2$, with $\partial(f^{-1}(V_0)) \cong \Sigma_g \times S^1$. Let $\nu(s_i)$ be a regular neighborhood of the arc s_i. Enlarging V_0 to include the critical value b_1, it can be shown that $f^{-1}(V_0 \cup \nu(s_1) \cup V_1)$ is diffeomorphic to $\Sigma_g \times D^2$ with a 2-handle h_1 attached along a circle γ_1 contained in a fiber $\Sigma_g \times pt \subset \Sigma_g \times S^1$. Moreover, Condition (2) in the definition of a Lefschetz fibration requires that h_1 is attached with a framing -1 relative to the natural framing on γ_1 inherited from the product structure of $\partial(f^{-1}(V_0))$. γ_1 is called a vanishing cycle. In addition, $\partial((\Sigma_g \times D^2) \cup h_1)$ is diffeomorphic to a Σ_g-bundle over S^1 whose monodromy is given by $D(\gamma_1)$,

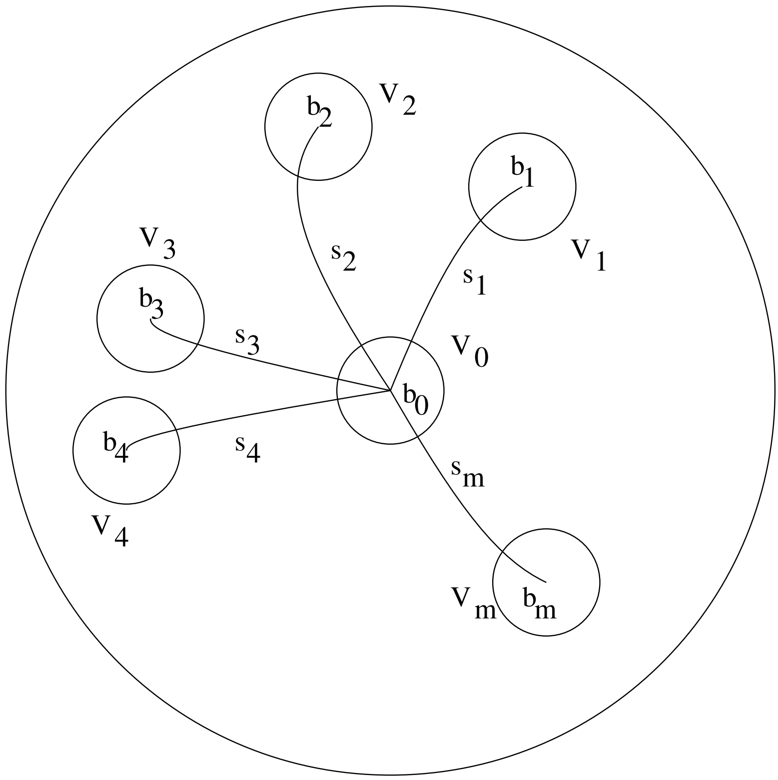

Figure 1. Fibration over the disk.

a positive Dehn twist about γ_1. Continuing counterclockwise about b_0, we add the remaining critical values to our description, yielding that

$$M_0 \cong f^{-1}\left(V_0 \cup \left(\bigcup_{i=1}^{m} \nu(s_i)\right) \cup \left(\bigcup_{i=1}^{m} V_i\right)\right)$$

is diffeomorphic to $(\Sigma_g \times D^2) \cup (\bigcup_{i=1}^{m} h_i)$, where each h_i is a 2-handle attached along a vanishing cycle γ_i in a Σ_g-fiber in $\Sigma_g \times S^1$ with relative framing -1. This handle attaching procedure will be explained in more detail in Section 2. Furthermore,

$$\partial M_0 \cong \partial\left((\Sigma_g \times D^2) \cup \left(\bigcup_{i=1}^{m} h_i\right)\right)$$

is a Σ_g-bundle over S^1 with monodromy given by the composition $D(\gamma_m) \cdots D(\gamma_1)$. We will refer to the cyclically ordered collection $(D(\gamma_1), \ldots, D(\gamma_m))$ (or the product $D(\gamma_m) \cdots D(\gamma_1)$) as the *global monodromy* of this fibration.

A Lefschetz fibration $f : M \to D^2$ does not completely determine the ordered collection $(D(\gamma_1), \dots, D(\gamma_m))$. Aside from the cyclic permutations and being able to conjugate all elements by a fixed arbitrary element of Γ_g, different choices of $\{s_i\}$ will give different monodromies. Given two choices of $\{s_i\}$, it is possible to get between them by a sequence of moves and their inverses. These moves which are called *elementary transformations*, can be thought of as the Lefschetz analog of handle slides in Morse theory. Each move interchanges the corresponding vanishing cycles, and also acts on one of the two cycles by the monodromy of the other. Equivalently, the pair of Dehn twists $(D(\gamma_i), D(\gamma_{i+1}))$ is replaced by $(D(\gamma_{i+1}), D(\gamma_{i+1})D(\gamma_i)D(\gamma_{i+1})^{-1})$. Thus, two relatively minimal Lefschetz fibrations over D^2 will be equivalent if and only if it is possible to get between the corresponding ordered collections of monodromies by a sequence of elementary transformations (and their inverses), together with an inner automorphism of $\mathcal{M}_g$ (cf. [**GS**]).

We can extend this description to Lefschetz fibrations over S^2 as follows:

Assume that $f : M \to S^2$ is a smooth genus g Lefschetz fibration. Let $M_0 = M - \nu(f^{-1}(b))$, where $\nu(f^{-1}(b)) \cong \Sigma_g \times D^2$ denotes a regular neighborhood of a nonsingular fiber $f^{-1}(b)$. Then $f|_{M_0} : M_0 \to D^2$ is a smooth Lefschetz fibration. If $(D(\gamma_1), \dots, D(\gamma_m))$ is the global monodromy of the fibration $f|_{M_0} : M_0 \to D^2$, then $D(\gamma_m) \cdots D(\gamma_1)$ is isotopic to the identity since also $\partial M_0 \cong \Sigma_g \times S^1$. Finally, to extend our description of M_0 to M, we reattach $\Sigma_g \times D^2$ to $(\Sigma_g \times D^2) \cup (\bigcup_{i=1}^m h_i)$ via a Σ_g-fiber preserving map of the boundary. This extension is unique up to equivalence for $g \geq 2$ [**K**].

Definition. Let $f : M \to S^2$ be a smooth genus g Lefschetz fibration with global monodromy $(D(\gamma_1), \dots, D(\gamma_m))$. We will call $f : M \to S^2$ a *hyperelliptic Lefschetz fibration* of genus g iff there exists $h \in \mathcal{M}_g$ such that $hD(\gamma_i)h^{-1} \in \mathcal{H}_g$ for all i, $1 \leq i \leq m$.

Remark. All Lefschetz fibrations of genus one and genus two are hyperelliptic since $\mathcal{H}_g = \mathcal{M}_g$ for $g = 1, 2$.

1.3. Wall's non-additivity formula.

If two compact oriented 4-manifolds are glued by an orientation reversing diffeomorphism of their boundaries, then the signature of their union is the sum of their signatures. This is known as the Novikov additivity. But it is often desirable to consider the more general case of gluing: Along a common submanifold, which may itself have boundary, of the boundaries of the original manifolds. However, the Novikov additivity does not hold in this general case. Wall [**W**] derives a formula for the deviation from additivity in the general case, which is known as the Wall's nonadditivity formula.

We will give a specific case of his formula:

Let X_-, X_0, X_+ be 3-manifolds and Y_- and Y_+ be 4-manifolds such that
$\partial X_- = \partial X_0 = \partial X_+ = Z$,
$\partial Y_- = X_- \cup X_0$,

$\partial Y_+ = X_0 \cup X_+$;
write $Y = Y_- \cup Y_+$ and $X = X_- \cup X_0 \cup X_+$ (Figure 2).

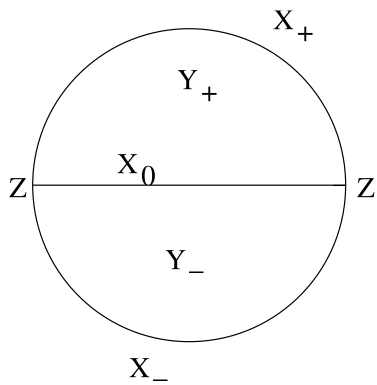

Figure 2.

Suppose that Y is oriented inducing orientations of Y_- and Y_+. Orient the rest so that
$\partial_*[Y_-] = [X_0] - [X_-]$,
$\partial_*[Y_+] = [X_+] - [X_0]$,
$\partial_*[X_-] = \partial_*[X_0] = \partial_*[X_+] = [Z]$.

Write $V = H_1(Z;\mathbb{R})$; let A, B and C be the kernels of the maps on first homology induced by the inclusions of Z in X_-,X_0 and X_+ respectively. Then $\dim A = \dim B = \dim C = \frac{(\dim V)}{2}$.

Let Φ denote the oriented intersection pairing in Z. Note that A,B and C are maximal isotropic subspaces for the intersection pairing Φ. Let $W = \frac{C\cap(A+B)}{(C\cap A)+(C\cap B)}$. Wall [**W**] defines a symmetric bilinear map $\Psi : W \times W \to \mathbb{R}$ as follows: The map $\Psi' : C \cap (A + B) \times C \cap (A + B) \to \mathbb{R}$ defined by $\Psi'(c, c') = \Phi(c, a')$ where $a' + b' + c' = 0$ for some $a' \in A$ and $b' \in B$ induces a well-defined bilinear map Ψ on W.

The signature of the symmetric bilinear map Ψ will be denoted by $\sigma(V; C, A, B)$.

We also denote the signature of a 4-manifold M as $\sigma(M)$ in the rest of this paper.

We are now ready to state Wall's formula:

Theorem 1 ([**W**])**.** *$\sigma(Y) = \sigma(Y_-) + \sigma(Y_+) - \sigma(V; C, A, B)$.*

1.4. Local signature formula.

The following theorem was proven by Matsumoto for $g = 1, 2$ using the fact that the cohomology class of Meyer's signature cocyle has finite order in the cohomology group $H^2(\mathcal{M}_g, \mathbb{Z})$. Recently, Endo proved the $g \geq 3$ case by observing the finiteness of the order of the cohomology class of the signature cocycle restricted to the hyperelliptic mapping class group $\mathcal{H}_g$.

Theorem 2 ([**M1**], [**M2**], [**E**]). *Let M be a 4-manifold which admits a hyperelliptic Lefschetz fibration of genus g over S^2. Let n and $s = \sum_{h=1}^{[\frac{g}{2}]} s_h$ be the numbers of nonseparating and separating vanishing cycles in the global monodromy of this fibration, respectively. Then*

$$\sigma(M) = -\frac{g+1}{2g+1}n + \sum_{h=1}^{[\frac{g}{2}]} \left(\frac{4h(g-h)}{2g+1} - 1 \right) s_h.$$

Remarks. (1) Here s_h denotes the number of separating vanishing cycles which separate the genus g surface into two surfaces one of which has genus h.

(2) This formula was reproven for the case $g = 2$ in [**Sm**].

2. Main theorems.

In this section we explain our main idea and establish the main theorems to develop an algorithm to compute the signature of a 4-manifold which admits a Lefschetz fibration over D^2 or S^2 using the global monodromy of this fibration.

Definition. Let X be a 4-manifold with boundary $\partial X \cong \Sigma_g \times I/(x,1) \sim (\phi(x), 0)$, where ϕ is a self-diffeomorphism of Σ_g. Let X' denote the resulting 4-manifold after attaching a 2-handle to X along a simple closed curve γ on $\Sigma_g \times \{pt\}$ with framing -1 (relative to the product framing). Then $\sigma(\phi, \gamma)$ is defined as $\sigma(X') - \sigma(X)$.

Theorem 3. *Let M be a 4-manifold which admits a genus g Lefschetz fibration over D^2 or S^2. Let $(D(\gamma_1), \ldots, D(\gamma_t))$ be the global monodromy of this fibration. Let $D(\gamma_0)$ denote the identity map. Then*

$$\sigma(M) = \Sigma_{i=1}^{t} \sigma(D(\gamma_{i-1}) \cdots D(\gamma_0), \gamma_i),$$

where $\sigma(D(\gamma_{i-1}) \cdots D(\gamma_1), \gamma_i) \in \{-1, 0, +1\}$ for all i, $1 \leq i \leq t$.

Proof. It suffices to prove the result for Lefschetz fibrations over D^2. (By Novikov additivity it extends to Lefschetz fibrations over S^2.) We use the handlebody description of M and Wall's formula as follows:

We start with a copy of $M_0 = \Sigma_g \times D^2$. We attach a 2-handle to M_0 along γ_1 with framing -1. Let M_1 denote the resulting manifold. Then ∂M_1 will have monodromy $D(\gamma_1)$, a positive Dehn twist about γ_1. Now we attach another 2-handle to M_1 along γ_2. Let M_2 denote the resulting manifold. Proceeding in this manner we get the manifolds $M_1, M_2, \ldots, M_t$.

We are going to apply Wall's formula at each step of this contruction to compute the signature of M. In order to apply Wall's formula we set up the following notation:

Take ϕ, X, γ and X' as in the definition above.

Let $\nu(\gamma)$ denote a regular neighborhood of γ in ∂X, and let i_* be the induced map on the homology by the inclusion of appropriate spaces.

Now we define Y_+, Y_-, X_+, X_0, X_-, Z in Wall's formula as follows:

$Y_- = D^2 \times D^2, \quad Y_+ = X,$
$\partial Y_- = \partial(D^2 \times D^2) = S^1 \times D^2 \cup D^2 \times S^1, \quad \partial Y_+ = \Sigma_g \times I/(x,1) \sim (\phi(x), 0),$
$X_0 = S^1 \times D^2 \cong \nu(\gamma), \quad X_- = D^2 \times S^1, \quad X_+ = \partial X - \mathring{\nu}(\gamma),$
$Z = S^1 \times S^1 \cong \partial\nu(\gamma) \cong \partial(\partial X - \mathring{\nu}(\gamma)).$

Hence,
$A = \mathrm{Ker}(i_* : H_1(S^1 \times S^1; \mathbb{R}) \to H_1(D^2 \times S^1; \mathbb{R})),$
$B = \mathrm{Ker}(i_* : H_1(S^1 \times S^1; \mathbb{R}) \to H_1(S^1 \times D^2; \mathbb{R})),$
$C = \mathrm{Ker}(i_* : H_1(\partial\nu(\gamma); \mathbb{R}) \to H_1(\partial X - \mathring{\nu}(\gamma); \mathbb{R})).$

Let l be the longitude $S^1 \times \{pt\}$ and m be the meridian $\{pt\} \times \partial D^2$ of $X_0 = S^1 \times D^2$. Then $A = \langle [l] \rangle$ and $B = \langle [m] \rangle$. We also know that C is a 1-dimensional subspace of

$$H_1(S^1 \times S^1; \mathbb{R}) = \langle [l], [m] \rangle \cong \mathbb{R}^2.$$

Let Φ be the intersection form on $Z = S^1 \times S^1$ and $W = \frac{C \cap (A+B)}{(C\cap A)+(C\cap B)} = \frac{C}{(C\cap A)+(C\cap B)}$. Hence $W = \{0\}$ if $C = A$ or $C = B$ and $W = C$ otherwise. Now assume that $C \neq A$ and $C \neq B$. Then $C = \langle c \rangle = \langle p[l] + q[m] \rangle$ for some $p, q \in \mathbb{R}$ and $\Psi(c,c) = \Phi(c, a')$ where $c + a' + b' = 0$ for some $a' \in A$ and $b' \in B$. (Ψ is the bilinear form in Wall's formula). Let $a' = -p[l]$ and $b' = -q[m]$. Then we have,

$$\Psi(c,c) = \Phi(c, -p[l]) = \Phi(p[l] + q[m], -p[l]) = -pq\Phi([m],[l]) = pq.$$

Therefore signature of Ψ is given by the sign of pq.

Hence by Wall's formula

$$\begin{aligned} \sigma(X') &= \sigma(X) + \sigma(D^2 \times D^2) - \sigma(\mathbb{R}^2; C, A, B) \\ &= \sigma(X) - \mathrm{signature}(\Psi) = \sigma(X) - \mathrm{sign}(pq). \end{aligned}$$

This proves the theorem by setting $X = M_i$ for $i = 1, 2, \ldots, t-1$. □

So the idea to compute the signature of a genus g Lefschetz fibration is very simple. For each 2-handle that we attach to $\Sigma_g \times D^2$ along a vanishing cycle, there is a corresponding *signature contribution* $\in \{-1, 0, +1\}$. Once we attach all the 2-handles, the sum of the signature contributions will be signature of the 4-manifold. The difficulty is to compute the signature contributions using the vanishing cycles (or more precisely using only the homology classes of the vanishing cycles). The following technical theorems will be helpful in computations.

Theorem 4. *In addition to the notation above, let* $\{a_1, b_1, a_2, b_2, \dots, a_g, b_g\}$ *be the standard basis for* $H_1(\Sigma_g; \mathbb{R})$. *(We will use the letters* a_i *and* b_i *also to denote the curves which represent the homology classes* a_i *and* b_i*, respectively, for* $1 \leq i \leq g$.*) Then:*

(1) *If* γ *is a nonseparating curve, then there exists a longitude* l' *and a meridian* m' *of* $\partial(\partial X - \overset{\circ}{\nu}(\gamma))$ *such that*

$$i_*[l'] = [\gamma] \in H_1(\partial X - \overset{\circ}{\nu}(\gamma); \mathbb{R})$$

$$i_*[m'] = \frac{e - \phi_*(e)}{e.[\gamma]} \in H_1(\partial X - \overset{\circ}{\nu}(\gamma); \mathbb{R})$$

for all $e \in \{a_1, b_1, a_2, b_2, \dots, a_g, b_g\}$*, where* $e.[\gamma] \neq 0$.

(2) *If* γ *is a separating curve, then* $\sigma(X') = \sigma(X) - 1$*, i.e.,* $\sigma(\phi, \gamma) = -1$.

Proof. We recall that ∂X is a mapping torus, i.e., $\partial X \cong \Sigma_g \times I/(x, 1) \sim (\phi(x), 0)$ and γ is a curve on a fiber $\Sigma_g \times \{pt\}$. We note that a regular neighborhood of γ in Σ_g is given by $\gamma \times I_1$. Hence a regular neighborhood of γ in ∂X is given by $\gamma \times I_1 \times I_2$ where I_2 is a small neighborhood of the $\{pt\}$ in $S^1 = I/(1 \sim 0)$. This neighborhood of γ is called the product neighborhood [**K**].

Now let us push off γ to the boundary of $\partial X - \overset{\circ}{\nu}(\gamma)$. Denote the push off of γ as l'. Moreover if we identify $I_1 \times I_2$ as D^2 and denote ∂D^2 as m', then $\{l', m'\}$ will be a longitude-meridian pair for $\partial(\partial X - \overset{\circ}{\nu}(\gamma))$. Then clearly

$$i_*[l'] = [\gamma] \in H_1(\partial X - \overset{\circ}{\nu}(\gamma); \mathbb{R}).$$

On the other hand, to find the image of m' we observe the following:

Assume that $e.[\gamma] = 1$ for some $e \in \{a_1, b_1, a_2, b_2, \dots, a_g, b_g\}$. Then we locally have the picture in Figure 3 in a neighborhood of the point where e and γ meet.

This proves that

$$i_*[m'] = e - \phi_*(e) \in H_1(\partial X - \overset{\circ}{\nu}(\gamma); \mathbb{R}).$$

Note that here we can deform $e - m'$ into $\phi_*(e)$ since the part of e which is not along m' lies outside of $\overset{\circ}{\nu}(\gamma)$.

Now assume that $e.[\gamma] = -1$ for some $e \in \{a_1, b_1, a_2, b_2, \dots, a_g, b_g\}$. Then we locally have a similar picture in a neighborhood of the point where e and γ meet, except for the orientations.

This proves that

$$i_*[m'] = \phi_*(e) - e \in H_1(\partial X - \overset{\circ}{\nu}(\gamma); \mathbb{R}).$$

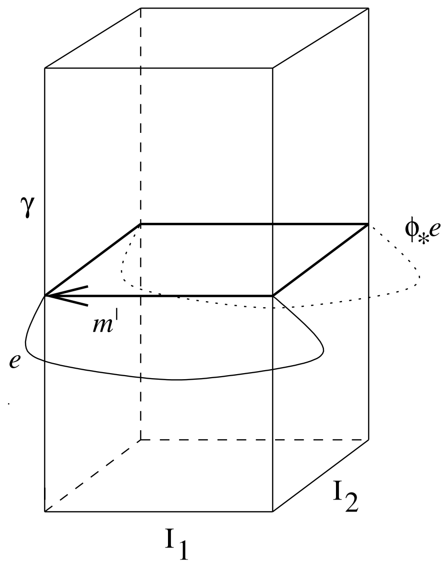

Figure 3.

Since these are local results it follows combining these two observations that

$$i_*[m'] = \frac{e - \phi_*(e)}{e.[\gamma]} \in H_1(\partial X - \mathring{\nu}(\gamma); \mathbb{R}).$$

To prove the second part of the theorem we note that if γ is a separating curve in Σ_g then it is homologically trivial. Thus $i_*[l'] = 0$. This implies that $\mathrm{Ker}(i_*) = \langle [l'] \rangle$.

Note that, in terms of the bases $\{[l], [m]\}$ of $H_1(\partial(S^1 \times D^2); \mathbb{R}) = H_1(S^1 \times S^1; \mathbb{R})$ and $\{[l'], [m']\}$ of $H_1(\partial(\partial X - \mathring{\nu}(\gamma)); \mathbb{R})$, attaching a 2-handle by -1 framing means that we identify $[l]$ with $[l'] - [m']$ and $[m]$ with $[m']$.

So if we transform the $\mathrm{Ker}(i_*)$ to the $\{[l], [m]\}$ plane we see that $\mathrm{Ker}\,(i_*) = C = W = \langle [l] + [m] \rangle$ which implies that $\sigma(X') = \sigma(X) - (+1)$ (cf. Theorem 3). □

Proposition 5. *We use the same notation as in Theorem* 4.

(1) *Let* $\gamma = a_i$ *for some* i, $1 \leq i \leq g$. *Then* $H_1(\partial X - \mathring{\nu}(a_i); \mathbb{R}) = \langle a_1, b_1, a_2, b_2, \ldots, a_g, b_g, b'_i, t \mid a_j = \phi_* a_j$ *for all* j, $b_j = \phi_* b_j$ *for all* $j \neq i$, $b'_i = \phi_* b_i \rangle$.
   Moreover $i_*[l'] = a_i$ *and* $i_*[m'] = b_i - b'_i$.

(2) *Let* $\gamma = b_i$ *for some* i, $1 \leq i \leq g$. *Then* $H_1(\partial X - \mathring{\nu}(b_i); \mathbb{R}) = \langle a_1, b_1, a_2, b_2, \ldots, a_g, b_g, a'_i, t \mid b_j = \phi_* b_j$ *for all* j, $a_j = \phi_* a_j$ *for all* $j \neq i$, $a'_i = \phi_* a_i \rangle$.
   Moreover $i_*[l'] = b_i$ *and* $i_*[m'] = a'_i - a_i$.

Proof. Assume that $\gamma = a_i$ for some i, $1 \le i \le g$. We first use Van-Kampen's theorem to compute $\pi_1(\partial X - \overset{\circ}{\nu}(a_i))$. Write $\partial X - \overset{\circ}{\nu}(a_i) = E_1 \cup E_2$ as follows: Let $E_1 = \Sigma_g \times [0, 1/2]$ and $E_2 = \Sigma_g \times [1/2, 1]$. Then glue $\Sigma_g \times \{1/2\} \subset E_1$ with $\Sigma_g \times \{1/2\} \subset E_2$ by the identity map except a neighborhood of a_i, namely $a_i \times I \subset \Sigma_g$. Denote the result as E'. By a trivial calculation we get the following presentation:

$$\pi_1(E') = \left\langle a_1, b_1, a_2, b_2, \ldots, a_g, b_g, b'_i \mid \prod_{j=1}^{g}[a_j, b_j],\ [a_i, b'_i]\prod_{j\neq i}[a_j, b_j] \right\rangle.$$

Finally we Abelianize this presentation after gluing $\Sigma_g \times \{0\} \subset E_1$ with $\Sigma_g \times \{1\} \subset E_2$ using the map ϕ to get $\partial X - \overset{\circ}{\nu}(a_i)$.

$i_*[l'] = a_i$ and $i_*[m'] = b_i - b'_i$ follows from Theorem 4 because a_i intersects b_j only once iff $i = j$.

Second part is obtained similarly. □

3. The algorithm and examples.

Suppose that we are attaching a 2-handle along a simple closed curve γ to a 4-manifold M with boundary $\partial M \cong \Sigma_g \times I/(x, 1) \sim (\phi(x), 0)$, where ϕ is a self-diffeomorphism of Σ_g. To compute the signature contribution of this handle we first compute $C = \ker i_*$ (cf. Wall's formula) using Theorem 4 and Proposition 5 and then apply Theorem 3. The signature contribution of a 2-handle will depend on the action of ϕ on $H_1(\Sigma_g)$ and the homology class $[\gamma] \in H_1(\Sigma_g)$.

3.1. Genus 1.

To illustrate how one can develop an algorithm using our main theorems to calculate the signatures of smooth Lefschetz fibrations, we will give the details of our computation to obtain the well-known result $\sigma(E(1)) = -8$ for the elliptic surface $E(1)$.

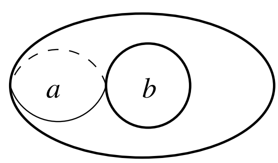

Figure 4.

The global monodromy of $E(1)$ is given by the sequence $(\alpha, \beta)^6$ of 12 Dehn twists where $\alpha = D(a)$ and $\beta = D(b)$ denote the positive Dehn twists about the curves a and b, respectively (Figure 4).

To build up $E(1)$, we start with a copy of $T^2 \times D^2$ and glue 2-handles along the vanishing cycles a and b in an alternating fashion. (We will use the letters a and b also to denote the homology classes of the curves a and b, respectively.)

Let ϕ denote the monodromy of the boundary of the 4-manifold before we attach a 2-handle.

We take A, B and C as in the proof of Theorem 3 and we apply Proposition 5 to compute $i_*[l']$ and $i_*[m']$. Note that we identify $[l]$ with $[l'] - [m']$ and $[m]$ with $[m']$ as in the proof of Theorem 4.

$\phi = \textit{identity}$, attach the first handle along a,
$i_*[l'] = a$ and $i_*[m'] = b - b' = b - \phi_* b = b - b = 0$,
$C = \langle [m'] \rangle = \langle [m] \rangle = B$ and therefore $\sigma(id, a) = 0$.

$\phi = \alpha$, attach the second handle along b,
$i_*[l'] = b$ and $i_*[m'] = a' - a = \phi_* a - a = 0$,
$C = \langle [m'] \rangle = \langle [m] \rangle = B$ and therefore $\sigma(\alpha, b) = 0$.

$\phi = \beta\alpha$, attach the third handle along a,
$i_*[l'] = a$ and $i_*[m'] = b - b' = b - \phi_* b = b - a = -a$ since $a = \phi_* a = a - b$,
$C = \langle [m'] + [l'] \rangle = \langle 2[m] + [l] \rangle$ and therefore $\sigma(\beta\alpha, a) = -1$.

$\phi = \alpha\beta\alpha$, attach the fourth handle along b,
$i_*[l'] = b$ and $i_*[m'] = a' - a = \phi_* a - a = -b - a = -2b$ since $b = \phi_* b = a$,
$C = \langle [m'] + 2[l'] \rangle = \langle 3[m] + 2[l] \rangle$ and therefore $\sigma(\alpha\beta\alpha, b) = -1$.

$\phi = \beta\alpha\beta\alpha$, attach the fifth handle along a,
$i_*[l'] = a$ and $i_*[m'] = b - b' = b - \phi_* b = b - (a - b) = 2b - a = 3b$ since $a = \phi_* a = -b$,
$C = \langle [m'] + 3[l'] \rangle = \langle 4[m] + 3[l] \rangle$ and therefore $\sigma(\beta\alpha\beta\alpha, a) = -1$.

$\phi = \alpha\beta\alpha\beta\alpha$, attach the sixth handle along b,
$i_*[l'] = b = 0$ and $i_*[m'] = a' - a = \phi_* a - a = -a - b - a = -2a - b = -2a$ since $b = \phi_* b = -b$,
$C = \langle [l'] \rangle = \langle [m] + [l] \rangle$ and therefore $\sigma(\alpha\beta\alpha\beta\alpha, b) = -1$.

$\phi = \beta\alpha\beta\alpha\beta\alpha$, attach the seventh handle along a,
$i_*[l'] = a = 0$ since $a = \phi_* a = -a$ and $i_*[m'] = b - b' = b - \phi_* b = b - (-b) = 2b$,
$C = \langle [l'] \rangle = \langle [m] + [l] \rangle$ and therefore $\sigma(\beta\alpha\beta\alpha\beta\alpha, a) = -1$.

$\phi = \alpha\beta\alpha\beta\alpha\beta\alpha$, attach the eighth handle along b,
$i_*[l'] = b$ and $i_*[m'] = a' - a = \phi_* a - a = -a - a = -2a = 4b$ since $b = \phi_* b = -a - b$,
$C = \langle -[m'] + 4[l'] \rangle = \langle 3[m] + 4[l] \rangle$ and therefore $\sigma(\alpha\beta\alpha\beta\alpha\beta\alpha, b) = -1$.

$\phi = \beta\alpha\beta\alpha\beta\alpha\beta\alpha$, attach the ninth handle along a,
$i_*[l'] = a$ and $i_*[m'] = b - b' = b - \phi_* b = b + a = 2a + a = 3a$ since $a = \phi_* a = -a + b$,
$C = \langle -[m'] + 3[l'] \rangle = \langle 2[m] + 3[l] \rangle$ and therefore $\sigma(\beta\alpha\beta\alpha\beta\alpha\beta\alpha, a) = -1$.

$\phi = \alpha\beta\alpha\beta\alpha\beta\alpha\beta\alpha$, attach the tenth handle along b,
$i_*[l'] = b$ and $i_*[m'] = a' - a = \phi_* a - a = b - a = 2b$ since $b = \phi_* b = -a$,
$C = \langle -[m'] + 2[l'] \rangle = \langle [m] + [l] \rangle$ and therefore $\sigma(\alpha\beta\alpha\beta\alpha\beta\alpha\beta\alpha, b) = -1$.

$\phi = \beta\alpha\beta\alpha\beta\alpha\beta\alpha\beta\alpha$, attach the eleventh handle along a,
$i_*[l'] = a$ and $i_*[m'] = b - b' = b - \phi_* b = b - (-a + b) = a$ since $a = \phi_* a = b$,
$C = \langle -[m'] + [l'] \rangle = \langle [l] \rangle = A$ and therefore $\sigma(\beta\alpha\beta\alpha\beta\alpha\beta\alpha\beta\alpha, a) = 0$.

$\phi = \alpha\beta\alpha\beta\alpha\beta\alpha\beta\alpha\beta\alpha$, attach the twelfth handle along b,
$i_*[l'] = b$ and $i_*[m'] = a' - a = \phi_* a - a = (a + b) - a = b$,
$C = \langle -[m'] + [l'] \rangle = \langle [l] \rangle = A$ and therefore $\sigma(\alpha\beta\alpha\beta\alpha\beta\alpha\beta\alpha\beta\alpha, b) = 0$.

Therefore by Theorem 3

$$\begin{aligned}\sigma(E(1)) &= \sigma(id, a) + \sigma(\alpha, b) + \sigma(\beta\alpha, a) + \cdots + \sigma(\alpha\beta\alpha\beta\alpha\beta\alpha\beta\alpha\beta\alpha, b) \\ &= 0 + 0 - (\underbrace{1 + \cdots + 1}_{8}) + 0 + 0 = -8.\end{aligned}$$

3.2. Genus 2.

We developed a Mathematica program to compute the signature of a 4-manifold which admits a genus two Lefschetz fibration over D^2 or S^2 whose global monodromy is given by any finite sequence of positive Dehn twists $D(c_1), D(c_2), \ldots, D(c_5)$, where $c_1, \ldots, c_5$ are the curves indicated in Figure 5

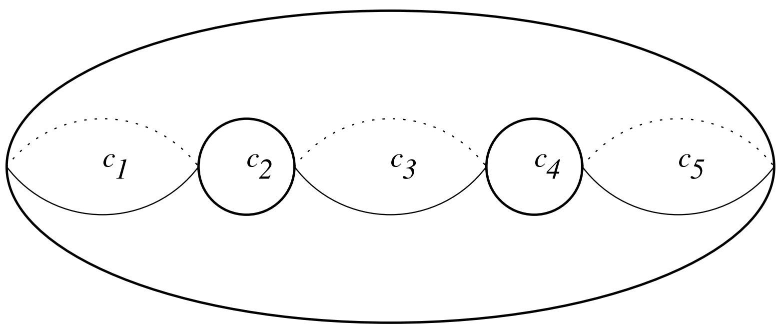

Figure 5.

Let ζ_i denote $D(c_i)$, $1 \leq i \leq 5$.

It was shown in [**M2**] that $\mathbb{C}P^2 \# 13\overline{\mathbb{C}P^2}$ admits a smooth Lefschetz fibration of genus two with global monodromy $(\zeta_1, \zeta_2, \zeta_3, \zeta_4, \zeta_5, \zeta_5, \zeta_4, \zeta_3, \zeta_2, \zeta_1)^2$. We computed the signature of the total space as

$$\begin{aligned}&\sigma((\zeta_1, \zeta_2, \zeta_3, \zeta_4, \zeta_5, \zeta_5, \zeta_4, \zeta_3, \zeta_2, \zeta_1)^2) \\ &= 0 + 0 + 0 + 0 - (\underbrace{1 + \cdots + 1}_{12}) + 0 + 0 + 0 + 0 \\ &= -12.\end{aligned}$$

One can also compute the signature of the total space starting from a cyclic permutation of the word above as follows.

$\sigma(\zeta_2, \zeta_3, \zeta_4, \zeta_5, \zeta_5, \zeta_4, \zeta_3, \zeta_2, \zeta_1, \zeta_1, \zeta_2, \zeta_3, \zeta_4, \zeta_5, \zeta_5, \zeta_4, \zeta_3, \zeta_2, \zeta_1, \zeta_1) =$

$$0 + 0 + 0 + 0 - (\underbrace{1 + \cdots + 1}_{11}) + 0 + 0 + 0 - 1 + 0 = -12.$$

Similarly a genus two Lefschetz fibration on $K3\#2\overline{\mathbb{C}P^2}$ is given in [**M2**] with the global monodromy $(\zeta_1, \zeta_2, \zeta_3, \zeta_4, \zeta_5)^6$.

We computed that $\sigma((\zeta_1, \zeta_2, \zeta_3, \zeta_4, \zeta_5)^6) =$

$$0+0+0+0-(\underbrace{1 + \cdots + 1}_{9})+0+0+0+0-(\underbrace{1 + \cdots + 1}_{9})+0+0+0+0 = -18.$$

Matsumoto [**M2**] also shows that $(\zeta_1, \zeta_2, \zeta_3, \zeta_4)^5$ is the hyperelliptic involution in $\mathcal{M}_2$, inducing the relation $(\zeta_1, \zeta_2, \zeta_3, \zeta_4)^{10} = 1$.

We computed that $\sigma((\zeta_1, \zeta_2, \zeta_3, \zeta_4)^{10}) = 0 + 0 + 0 + 0 - (\underbrace{1 + \cdots + 1}_{8})+$

$$0 + 0 + 1 - (\underbrace{1 + \cdots + 1}_{10}) + 1 + 0 + 0 - (\underbrace{1 + \cdots + 1}_{8}) + 0 + 0 + 0 + 0 = -24.$$

Note that the 15th and the 26th 2-handle both contribute +1 to the signature. It is known that the total space of this fibration is homeomorphic but not diffeomorphic to $5\mathbb{C}P^2\#29\overline{\mathbb{C}P^2}$ (cf. [**M2**], [**F1**]).

As a final example we give the signature contributions of the singular fibers in the genus two Lefschetz fibration of $S^2 \times T^2\#4\overline{\mathbb{C}P^2}$ given in [**M2**].

$$\sigma(S^2 \times T^2\#4\overline{\mathbb{C}P^2}) = 0 - 1 - 0 - 1 - 1 - 1 + 0 + 0 = -4.$$

Remark. One can indeed check these numbers using Matsumoto's local signature formula or using the fact that $\sigma(X\#\overline{\mathbb{C}P^2}) = \sigma(X) - 1$, for a 4-manifold X.

3.3. Genus 3.

Let $\mathbb{F}_2$ denote the Hirzebruch surface, the holomorphic $\mathbb{C}P^1$ bundle over $\mathbb{C}P^1$ with a holomorphic section s_1 of self intersection -2. $\mathbb{F}_2$ also admits a disjoint holomorphic section s_2 of self intersection 2. Let X be the two-fold cover of $\mathbb{F}_2$, branced over the disjoint union of a smooth curve in $|7s_1|$ and s_2. Then X admits a holomorphic Lefschetz fibration $X \to \mathbb{C}P^1$ of genus three obtained by composing the branched cover map with the bundle map $\mathbb{F}_2 \to \mathbb{C}P^1$. In [**F2**], Fuller gives the global monodromy of this fibration as

$$(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5, \eta_6)^{14}.$$

Here $\eta_1, \eta_2, \ldots, \eta_9$ denote the positive Dehn twists about the curves $d_1, d_2, \ldots, d_9$ indicated as in Figure 6.

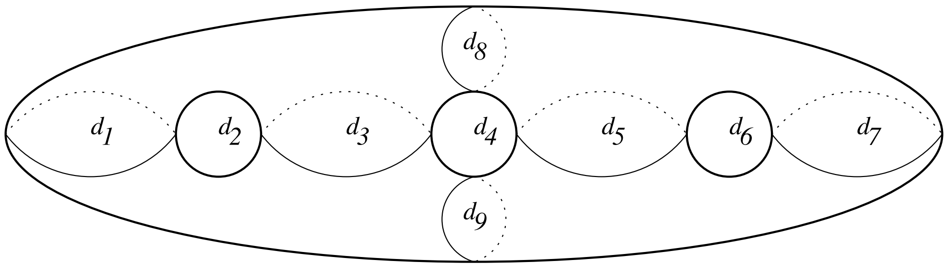

Figure 6.

We computed the signature of this genus three Lefschetz fibrations over S^2, using our Mathematica program. (The program is available upon request.)

$$\sigma((\eta_1,\eta_2,\eta_3,\eta_4,\eta_5,\eta_6)^{14}) =$$

$$\underbrace{0+\cdots+0}_{6}-(\underbrace{1+\cdots+1}_{8})+\underbrace{0+\cdots+0}_{4}-(\underbrace{1+\cdots+1}_{8})+\underbrace{0+\cdots+0}_{4}$$
$$-1-1+0+0+1-(\underbrace{1+\cdots+1}_{14})+1+0+0-1-1+$$
$$\underbrace{0+\cdots+0}_{4}-(\underbrace{1+\cdots+1}_{8})+\underbrace{0+\cdots+0}_{4}-(\underbrace{1+\cdots+1}_{8})+\underbrace{0+\cdots+0}_{6}$$
$$= -48.$$

Fuller (cf. [**F3**]) also derives the following word in $\mathcal{M}_3$.

$$(\eta_8,\eta_9,\eta_4,\eta_3,\eta_2,\eta_1,\eta_5,\eta_4,\eta_3,\eta_2,\eta_6,\eta_5,\eta_4,\eta_3,(\eta_1,\eta_2,\eta_3,\eta_4,\eta_5,\eta_6)^{10}).$$

We computed that

$$\sigma((\eta_8,\eta_9,\eta_4,\eta_3,\eta_2,\eta_1,\eta_5,\eta_4,\eta_3,\eta_2,\eta_6,\eta_5,\eta_4,\eta_3,(\eta_1,\eta_2,\eta_3,\eta_4,\eta_5,\eta_6)^{10}))$$
$$= -42.$$

The four manifold with the global monodromy given as above is not very familiar and it is our only example where we can not use any other method than ours to calculate the signature. The Lefschetz fibration is not hyperelliptic, for example, otherwise the local signature formula [**E**] would yield $\sigma = 74(-4/7)$ which is not an integer! It is not known whether this fibration is holomorphic or not.

Corollary 6. *There exist two genus three Lefschetz fibrations with the same Euler characteristic but having different signatures.*

Proof. Let M_1 and M_2 be the 4-manifolds with global monodromies

$$(\eta_1,\eta_2,\eta_3,\eta_4,\eta_5,\eta_6)^{14}$$

and

$$(\eta_8, \eta_9, \eta_4, \eta_3, \eta_2, \eta_1, \eta_5, \eta_4, \eta_3, \eta_2, \eta_6, \eta_5, \eta_4, \eta_3, (\eta_1, \eta_2, \eta_3, \eta_4, \eta_5, \eta_6)^{10})$$

respectively.

Let $\#_n M$ denote the n-fold fiber sum of M by itself. Then $e(\#_{37}M_1) = e(\#_{42}M_2)$ but $\sigma(\#_{37}M_1) = -1776$ and $\sigma(\#_{42}M_2) = -1764$. □

Remark. Following the language in [**Sm**], we say that two Lefschetz fibrations are combinatorially equivalent if they have the same fiber genus and the same number of each conjugacy type of singular fibers. The corollary above shows that signature is not an invariant of the combinatorial equivalence class of a Lefschetz fibration. Moreover there is not necessarily a hyperelliptic Lefschetz fibration in each combinatorial equivalence class.

4. Some applications.

First we give an immediate application of Theorems 3 and 4.

Corollary 7. *Let M be a 4-manifold which admits a genus g Lefschetz fibration over D^2 or S^2. Let n and s be the numbers of nonseparating and separating vanishing cycles in the global monodromy of this fibration, respectively. Then $\sigma(M) \leq n - s$.*

Proof. Suppose that we build up the 4-manifold M from $\Sigma_g \times D^2$ by attaching 2-handles. By Theorem 4, every time we attach a 2-handle along a separating curve the signature of the resulting 4-manifold will be one less than the signature of the 4-manifold before we attach the 2-handle. Thus Theorem 3 implies the upper bound $n - s$ on the signature. □

Remark. Define $c_1^2(M) = 3\sigma(M) + 2\chi(M)$ and $\chi_h(M) = \frac{1}{4}(\sigma(M) + \chi(M))$ for a closed symplectic 4-manifold M. Note that $\sigma(M) \leq n + s = \chi(M) + 4g - 4$ trivially implies $c_1^2 \leq 10\chi_h + 2g - 2$ for a genus g Lefschetz fibration over S^2.

Corollary 8. *There is no hyperelliptic Lefschetz fibration* (*of any genus*) *over S^2 with only reducible singular fibers.* (*Here reducible means that the local monodromy corresponding to the singular fiber is a Dehn twist about a separating curve.*)

Remark. In particular, this proves that a product of positive Dehn twists about separating curves can not be equal to the identity in $\mathcal{M}_2$, which is a result of Mess [**Me**].

Proof. Let M be a 4-manifold which admits a Lefschetz fibration of genus g over S^2 with global monodromy $(D(\gamma_1), \dots, D(\gamma_s))$, where $s = \sum_{h=1}^{[\frac{g}{2}]} s_h$ and γ_i is separating for each i, $1 \leq i \leq s$. Then, by the local signature formula,

$$\sigma(M) = \begin{cases} \sum_{h=1}^{[\frac{g}{2}]} \left(\frac{4h(g-h)}{2g+1} - 1\right) s_h \geq 0 & \text{if } g \geq 3 \\ -s/5 & \text{if } g = 2. \end{cases}$$

But on the other hand $\sigma(M) = -s$ according to Theorem 4. Hence $s = 0$. (This is trivially true for $g = 1$ since any vanishing cycle on a torus is nonseparating.) This proves the desired result since we assume (by definition) that there exists at least one singular fiber in each Lefschetz fibration. □

Next we combine our results with the local signature formula for the hyperelliptic Lefschetz fibrations to give an upper bound for the signatures of these fibrations.

Corollary 9. *Let M be a 4-manifold which admits a hyperelliptic Lefschetz fibration of genus g over S^2. Let n and s be the numbers of nonseparating and separating vanishing cycles in the global monodromy of this fibration, respectively. Then $\sigma(M) \leq n - s - 4$.*

Remark. This inequality is not necessarily sharp.

Proof. We first note that we can improve the inequality

$$\sigma(M) \leq n - s$$

given in Corollary 7 to

$$\sigma(M) \leq n - s - 1$$

for hyperelliptic Lefschetz fibrations as follows:

Suppose that we attach the first 2-handle along a nonseparating curve. We can always assume this because $n \geq 1$ (since we proved in Corollary 8 that $n \neq 0$) and we can cyclically permute the vanishing cycles in the global monodromy of a Lefschetz fibration. Moreover we can easily show that if we start attaching handles along a nonseparating curve then the signature of the resulting 4-manifold (after attaching the very first handle) will be the same as $\sigma(\Sigma_2 \times D^2)$, which is zero.

Next note that $\sigma(M) \leq n - s - 1$ is equivalent to

$$4 \sum_{h=1}^{[\frac{g}{2}]} h(g-h)s_h \leq (3g+2)n - (2g+1)$$

using the local signature formula.

Assume that g is odd. Endo [**E**] proves that

$$n + 4 \sum_{h=1}^{[\frac{g}{2}]} h(2h+1)s_h \equiv 0 \pmod{4(2g+1)}.$$

Hence

$$n = 4c(2g+1) - 4\sum_{h=1}^{[\frac{g}{2}]} h(2h+1)s_h$$

for some integer c. Substituting into the inequality above (and dividing by 4) we get

$$\sum_{h=1}^{[\frac{g}{2}]} h(g-h)s_h \leq (3g+2)\left[c(2g+1) - \sum_{h=1}^{[\frac{g}{2}]} h(2h+1)s_h\right] - \frac{1}{4}(2g+1).$$

Hence

$$\sum_{h=1}^{[\frac{g}{2}]} h(g-h)s_h \leq (3g+2)\left[c(2g+1) - \sum_{h=1}^{[\frac{g}{2}]} h(2h+1)s_h\right] - \frac{1}{4}(2g+2)$$

since $2g+1 \equiv 3 \pmod 4$.

But this inequality, in turn, implies that

$$4\sum_{h=1}^{[\frac{g}{2}]} h(g-h)s_h \leq (3g+2)n - (2g+2)$$

which is equivalent to

$$\sigma(M) \leq n - s - 1 - \frac{1}{2g+1}.$$

Since $\sigma(M)$ is an integer,

$$\sigma(M) \leq n - s - 2.$$

Iterating the same argument, we obtain

$$\sigma(M) \leq n - s - 4.$$

(We use $2(2g+1) \equiv 2 \pmod 4$ and $3(2g+1) \equiv 1 \pmod 4$.)

Similarly, if g is even, then one can use the corresponding result by Endo:

$$n + 4\sum_{h=1}^{[\frac{g}{2}]} h(2h+1)s_h \equiv 0 \pmod{2(2g+1)}.$$

(Note that $2(3g+2)(2g+1) \equiv 0 \pmod 4$, if g is even.) □

The following is a result concerning the *geography* of the hyperelliptic Lefschetz fibrations, which follows easily from Corollary 9.

Corollary 10. (1) *The total space of a genus two Lefschetz fibration over S^2 satisfies*

$$c_1^2 \leq 6\chi_h - 3.$$

(2) *The number of singular fibers in a genus two Lefschetz fibration over* S^2 *can not be equal to* $5, 6, 11$ *or* 12 *and in particular, the minimal number of singular fibers in a genus two Lefschetz fibration over* S^2 *is* 7 *or* 8.

(3) *The total space of a genus three hyperelliptic Lefschetz fibration over* S^2 *satisfies*

$$c_1^2 \leq 7.25\chi_h - 2.75.$$

Remarks. (1) In particular, the signature of a smooth 4-manifold which admits a hyperelliptic Lefschetz fibration of genus $g \leq 3$ over S^2 is negative.

(2) Similar inequalities can be obtained for genus $g \geq 4$ hyperelliptic Lefschetz fibrations over S^2.

5. Final remark.

Given a product of positive Dehn twists in the mapping class group of a genus g surface, we can construct a symplectic 4-manifold which admits a Lefschetz fibration over D^2, as we have studied in this paper. A natural generalization is to allow negative Dehn twists also. These fibrations are called *achiral* Lefschetz fibrations. Our technique clearly extends to compute the signatures of these fibrations.

References

[BH] J.Birman and H.Hilden, *On mapping class groups of closed surfaces as covering spaces,* Advances in the Theory of Riemann Surfaces, Annals of Math. Studies, **66**, Princeton Univ. Press, (1971), 81-115, Zbl 217.48602.

[D] S. Donaldson, *Lefschetz fibrations in symplectic geometry,* Doc. Math. J. DMV., Extra Volume, **ICMII** (1998), 309-314, MR 99i:57044, Zbl 909.53018.

[E] H. Endo, *Meyer's signature cocyle and hyperelliptic fibrations,* Math. Ann., **316** (2000), 237-257, MR 2001b:57047, Zbl 0948.57013.

[F1] T. Fuller, *Diffeomorphism types of genus* 2 *Lefschetz fibrations,* Math. Ann., **311**(**1**) (1998), 163–176, MR 99f:57035, Zbl 905.57014.

[F2] ———, *Hyperelliptic Lefschetz fibrations and branched covering spaces,* Pacific J. Math., **196** (2000), 369-393, CMP 1 800 582.

[F3] ———, private communication.

[GS] R. Gompf and A. Stipsicz, *An introduction to* 4*-manifolds and Kirby calculus,* AMS Graduate Studies in Mathematics, **20** (1999).

[K] A. Kas, *On the handlebody decomposition associated to a Lefschetz fibration*, Pacific J. Math., **89** (1980), 89-104, MR 82f:57012, Zbl 457.14011.

[L] T-J. Li, *Symplectic Parshin-Arakelov inequality,* Internat. Math. Res. Notices, **18** (2000), 941-954, CMP 1 792 283.

[M1] Y. Matsumoto, *On* 4*-manifolds fibered by tori* II, Proc. Japan Acad., **59A** (1983), 100-103, MR 84j:57010b, Zbl 532.55020.

[M2] ______, *Lefschetz fibrations of genus two - a topological approach,* Proceedings of the 37th Taniguchi Symposium on Topology and Teichmüller Spaces, ed. Sadayoshi Kojima et al., World Scientific (1996), 123-148, MR 2000h:14038, Zbl 921.57006.

[Me] G. Mess, *The Torelli groups for genus* 2 *and* 3 *surfaces*, Topology, **31**(**4**) (1992), 775-790, MR 93k:57003, Zbl 772.57025.

[Sm] I. Smith, *Lefschetz fibrations and the Hodge bundle*, Geom. Topol., **3** (1999), 211-233, MR 2000j:57059, Zbl 929.53047.

[St1] A. Stipsicz, *A note on the geography of symplectic manifolds*, Turkish J. Math., **20** (1996), 135-139, MR 97m:57035, Zbl 876.57039.

[St2] ______, *Erratum to: Chern numbers of certain Lefschetz fibrations,* Proc. Amer. Math. Soc., **128**(**9**) (2000), 2833-2834, CMP 1 766 607.

[St3] ______, *On the number of vanishing cycles in Lefschetz fibrations,* Math. Res. Lett., **6** (1999), 449-456, MR 2000g:57046, Zbl 955.57026.

[W] C.T.C. Wall, *Non-additivity of the signature,* Inventiones Math., **7** (1969), 269-274, MR 39 #7615, Zbl 176.21501.

Received January 8, 2000 and revised March 7, 2000.

Department of Mathematics
Michigan State University
East Lansing, MI 48824-1028
E-mail address: bozbagci@math.msu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 202, No. 1, 2002

EFFECTIVE DIVISOR CLASSES ON A RULED SURFACE

Jeff Rosoff

The Neron-Severi group of divisor classes modulo algebraic equivalence on a smooth algebraic surface is often not difficult to calculate, and has classically been studied as one of the fundamental invariants of the surface. A more difficult problem is the determination of those divisor classes which can be represented by effective divisors; these divisor classes form a monoid contained in the Neron-Severi group. Despite the finite generation of the whole Neron-Severi group, the monoid of effective divisor classes may or may not be finitely generated, and the methods used to explicitly calculate this monoid seem to vary widely as one proceeds from one type of surface to another in the standard classification scheme (see Rosoff, 1980, 1981).

In this paper we shall use concrete vector bundle techniques to describe the monoid of effective divisor classes modulo algebraic equivalence on a complex ruled surface over a given base curve. We will find that, over a base curve of genus 0, the monoid of effective divisor classes is very simple, having two generators (which is perhaps to be expected), while for a ruled surface over a curve of genus 1, the monoid is more complicated, having either two or three generators. Over a base curve of genus 2 or greater, we will give necessary and sufficient conditions for a ruled surface to have its monoid of effective divisor classes finitely generated; these conditions point to the existence of many ruled surfaces over curves of higher genus for which finite generation fails.

0. Preliminaries on ruled surfaces.

Let C be a nonsingular complex curve, and let X be a ruled surface over C. Then ([**H1**, V.2.2, p. 370]) X can be written as $\mathbf{P}(E)$, the projectivization of a rank 2 vector bundle E on C. Moreover, for rank 2 bundles E and E', $\mathbf{P}(E) = \mathbf{P}(E')$ if and only if E and E' differ by a twisting with a line bundle.

Let X be a ruled surface over the base curve C, say $X = \mathbf{P}(E')$. Replacing E' by $E = E' \otimes L$ for some line bundle L chosen appropriately, we may assume that E is *normalized* so that:

(a) E admits a nontrivial global section, and,
(b) $E \otimes L$ admits no nontrivial section for any line bundle L of negative degree.

E is determined up to twisting with line bundles of degree 0. The choice of a (normalized) bundle E to represent X as the projectivization of a rank two bundle determines the linear equivalence class of a "standard" section C_0 of X over C, with $\mathcal{O}_X(C_0) = \mathcal{O}_{\mathbf{P}(E)}(1)$.

If E is a rank 2 bundle on C, then $\deg \Lambda^2(E) = \deg \Lambda^2(E \otimes L)$ for any line bundle L of degree 0, and hence is a well-defined invariant of $X = \mathbf{P}(E)$ for E normalized as above. Finally, the Neron-Severi group of X is a free group of rank 2, generated by the algebraic equivalence classes of the section C_0 and of any fiber F of the natural map $\pi : X \to C$. Denoting by $[D]$ the algebraic equivalence class of any divisor D, the intersection pairing on X is given by:

$$[C_0]^2 = \deg \Lambda^2(E), \quad [C_0] \bullet [F] = 1, \quad [F]^2 = 0.$$

1. Ruled surfaces with $[C_0]^2 \leq 0$.

Theorem 1. *Let C be a smooth curve and let $X = \mathbf{P}(E)$ be a ruled surface over C such that, with E normalized as above, $[C_0]^2 = \deg \Lambda^2(E) \leq 0$. Then the monoid of effective divisor classes on X is $\{a[C_0] + b[F] \mid a, b \geq 0\}$.*

Remark. Since any vector bundle over $\mathbf{P}^1$ splits into a direct sum of line bundles, this theorem applies to any ruled surface over a curve of genus 0; a normalized rank 2 bundle will be of the form $\mathcal{O} \oplus L$ for some line bundle L of degree ≤ 0. These are the "Hirzebruch surfaces".

Proof. Clearly any divisor class in the above set represents an effective divisor. Now, let D be an effective divisor on X, so that D can be written as $D = nC_0 + \sum D_i$ with $n \geq 0$ and D_i an irreducible curve other than C_0. Letting $[D_i] = a_i[C_0] + b_i[F]$, we have that $[D_i] \bullet [F] = a_i \geq 0$ and $[D_i] \bullet [C_0] = b_i + a_i[C_0]^2 \geq 0$, so $b_i \geq 0$. The result follows by additivity.

2. Ruled surfaces over curves of genus 1.

Theorem 2. *Let $X = \mathbf{P}(E)$ be a ruled surface over an elliptic curve C, with E normalized. Then the monoid of effective divisor classes on X is (finitely) generated by:*

(a) $[C_0]$ *and* $[F]$ *if* $\deg \Lambda^2(E) \leq 0$, *and*
(b) $[C_0]$, $[F]$ *and the anti-cannonical class* $2[C_0] - [F]$ *if* $\deg \Lambda^2(E) > 0$.

Proof. We may assume that $\deg \Lambda^2(E) \geq 1$, and by Nagata [**N**, pp. 191-96] we may assume that this degree is 1.

Since E has a nontrivial section and since $E \otimes \mathcal{O}(-p)$ has none for any $p \in C$, we have an exact sequence of bundles $0 \to \mathcal{O}_C \to E \to \mathcal{O}(p) \to 0$, with E a nontrivial extension of $\mathcal{O}(p)$ by $\mathcal{O}_C$ for some fixed $p \in C$. The nontriviality of the extension corresponds to E being normalized.

We will first show that the divisor class $2[C_0] - [F]$ contains an effective divisor, i.e., that the bundle $\mathcal{O}_X(2) \otimes \pi^*\mathcal{O}_C(-q)$ has nontrivial section for some $q \in C$. By [**H1**, 7.11, p. 162] $\pi_*\mathcal{O}_X(2) = S^2(E)$, the second symmetric power of E, so by the projection formula it suffices to show that $S^2(E) \otimes \mathcal{O}_C(-q)$ has a section for some $q \in C$.

The structure of the space of indecomposable bundles of a given rank and degree on an elliptic curve has been extensively studied by Atiyah [**A**, pp. 414-52], and we will appeal to his results to show that $S^2(E) = \sum \mathcal{O}(p_i)$ where the $\{p_i\}$ are the three nontrivial half-periods of C under the translation sending p to the origin. In Atiyah's notation $E = e_A(2,1)$ where A is the bundle $\mathcal{O}(p)$, and by [**A**, Th. 7, p. 434], $\check{E} = e_A(2,-1) = E \otimes \mathcal{O}(-p)$ [**A**, Th. 6, p. 433].

Thus, $E \otimes E \otimes \mathcal{O}(-p) = E \otimes \check{E} = \operatorname{End}(E) = \mathcal{O}_C \oplus \sum \mathcal{O}(p_i - p)$ with the $\{p_i\}$ as above using [**A**, Lemma 22, p. 439], and so $E \otimes E = \mathcal{O}(p) \oplus \sum \mathcal{O}(p_i)$. On the other hand, over $\mathbf{C}$, $E \otimes E = S^2(E) \oplus \Lambda^2(E) = S^2(E) \oplus \mathcal{O}(p)$. Thus $S^2(E) = \sum \mathcal{O}(p_i)$ and $S^2(E) \otimes \mathcal{O}(-q)$ has a section for $q \in \{p_i\}$.

Thus the monoid generated by the classes $[C_0], [F]$ and $2[C_0] - [F]$ is contained in the monoid of effective divisor classes on X. The reverse inclusion follows directly from [**H1**, V.2.21, p. 382].

3. Ruled surfaces over curves of higher genus.

For the reminder of this article, we will assume that our ruled surface X is the projectivization of a normalized rank 2 bundle E on the curve C with $g(C) \geq 1$ and, in view of Theorem 1, that $\deg \Lambda^2(E) \geq 1$.

By [**H1**, V.2.21, p. 382], if Y is an irreducible curve on X with $[Y] = a[C_0] + b[F]$, then Hurwitz's theorem applied to the desingularization of Y shows that either $a, b \geq 0$, or $a \geq 2$ and $b \geq -(a/2)\deg \Lambda^2(E)$. It follows from linearity that these are necessary conditions for an algebraic equivalence class to contain an effective divisor.

On the other hand, sufficient conditions are provided by the Riemann-Roch theorem: By Serre duality $\dim H^2(X, \mathcal{O}_X(aC_0 + bF)) = 0$ for $a \geq 0$, and Riemann-Roch on the surface X applied to the divisor class $[aC_0 + bF]$ shows that this class contains an effective divisor if $a \geq 0$ and $b \geq g - (a/2)\deg \Lambda^2(E)$.

Theorem 3. *Let $X = \mathbf{P}(E)$ be a ruled surface over a curve C with $g(C) \geq 1$, and $\deg \Lambda^2(E) \geq 1$ with E normalized. Then the monoid of effective divisor classes on X is finitely generated if and only if there is effective divisor class $a[C_0] + b[F]$ with $a \geq 2$ and $b = -(a/2)\deg \Lambda^2(E)$.*

Proof. Suppose that there is an effective divisor class as above, and let $A \geq 2$ be minimal such that the class $A[C_0] - (A/2)\deg \Lambda^2(E)[F]$ contains an effective divisor, say D_0. Let $S = \{a[C_0] + b[F] \mid 1 \leq a < A$ and $-(a/2)\deg \Lambda^2(E) \leq b < 0\}$; S contains a finite number of (not necessarily effective) divisor classes, say $[E_1] \cdots [E_k]$.

Let $\check{S} = \{[E_i] \in S \,|\, [E_i] + n[D_0]$ contains an effective divisor for some $n \geq 0\}$, and for each $[E_i] \in \check{S}$ let $D_i = E_i + n_i D_0$ with $n_i \geq 0$ minimal in the above regard. Then the monoid of effective divisor classes on X is (finitely) generated by the classes $[C_0]$, $[F]$ and the $[D_i]$, $i \geq 0$.

Conversely, suppose that there is no effective divisor class on X meeting the condition of the theorem. Then, for any finite collection of (nontrivial) effective divisors $D_1, \dots, D_k$ we have $[D_i] = a_i[C_0] + b_i[F]$ with $a_i \geq 0$ and $b_i > -(a_i/2)\deg \Lambda^2(E)$, i.e., for some $\varepsilon_i > 0$ $b_i \geq -(a_i/2)\deg \Lambda^2(E) + a_i\varepsilon_i$. For any nonnegative integers $c_1, \dots, c_k$ (not all 0) the linear combination $A[C_0] + B[F] = c_1[D_1] + \cdots + c_k[D_k]$ has $B \geq -(A/2)\deg \Lambda^2(E) + A\varepsilon$ for $\varepsilon = \min\{\varepsilon_i\}$, $\varepsilon > 0$.

Finally, select an integer a sufficiently large so that $g(C) < \varepsilon a$; making a even if necessary, let $b \in \mathbf{Z}$ be $b = g(C) - (a/2)\deg \Lambda^2(E)$. By our above observation on Riemann-Roch, the divisor class $a[C_0] + b[F]$ is an effective class, but cannot be expressed as a nonnegative linear combination of the classes $\{[D_i]\}$. Thus, if there is no effective class as in the theorem, the monoid of effective divisor classes on X is not finitely generated.

A geometric criterion for finite generation of the monoid of effective classes is given by:

Corollary. *The monoid of effective divisor classes on X is finitely generated if and only if there is a curve Y on X such that the projection $\pi : X \to C$ exhibits Y as an unramified n-fold cover of C for some $n \geq 2$.*

Proof. Indeed, the proof of [**H1**, V.2.21, p. 382] shows that an irreducible curve Y on X that is an unramified n-fold cover of C (for $n \geq 2$) must necessarily have $[Y] = n[C_0] - (n/2)\deg \Lambda^2(E)[F]$, and that any curve in a divisor class of this form (with n minimal and $n \geq 2$) is such a cover.

Remark. For any such curve Y on X as above, we must have $\dim H^0(X, \mathcal{O}_X(Y)) = 1$. To see this, note that if $D \in |Y|$, then $D \cap \pi^{-1}(u)$ consists of n distinct points for all $u \in C$. If there are two linearly independent sections s_i in $H^0(X, \mathcal{O}_X(Y))$, then as $Y^2 = 0$, we have a surjection $\Psi : X \to \mathbf{P}^1$ given by $\Psi(p) = [s_1(p), s_2(p)]$. Since $\Psi^{-1}(t) \in |Y|$ for $t \in \mathbf{P}^1$, $\Psi|F$ gives an unramified n-fold cover of $\mathbf{P}^1$ for a fiber F, which is impossible for $n > 1$.

Theorem 4. *Let X be the ruled surface $\mathbf{P}(E)$ over the curve C with $g(C) \geq 1$, with E normalized and with $\deg \Lambda^2(E) > 0$. Then finite generation of the monoid of effective divisor classes on X is equivalent to the existence* *of a sub-line bundle N of $S^n(E)$ with* $\deg(N) = (n/2)\deg \Lambda^2(E)$, *for some* $n > 1$.

Proof. [**H2**, 10.2, p. 51] shows that a multisection Y of $\mathbf{P}(E)$ of degree n over C corresponds to a sub-line bundle N of $S^n(E)$, and further that the requirement that $[Y] = n[C_0] - (n/2)\deg \Lambda^2(E)[F]$ forces $\deg(N) = (n/2)\deg \Lambda^2(E)$.

Remark. An interesting geometric proof of Theorem 3 can be obtained by considering the projective n-bundle $\mathbf{P}(S^n\check{E})$ over C, with projection ψ to C. Giving a multisection Y of $\mathbf{P}(E)$ of degree n over C can be viewed as giving a section σ of $\mathbf{P}(S^n\check{E})$ — this is really the same as giving local homogeneous equations defining Y on each fiber of π in $\mathbf{P}(E)$, with the fibers of ψ parameterizing homogeneous polynomials of degree n in two variables. There is a divisor Δ on $\mathbf{P}(S^n\check{E})$ whose intersection with each fiber of ψ corresponds to such local equations having zero discriminant; a linear algebra computation shows that $O(\Delta) = \mathcal{O}(2n-2) \otimes \psi^*(L)$ for some line bundle L on C of degree $n(n-1)\deg \Lambda^2(E)$. The requirement that Y be an unramified n-fold cover of C is that $\sigma(C)$ not meet Δ, so that $\deg \sigma^*(\Delta) = 0$, i.e., $\deg \sigma^*\mathcal{O}(1) = -(n/2)\deg \Lambda^2(E)$ using $n \geq 2$. Since $\sigma^*\mathcal{O}(1)$ is a quotient of $S^n(\check{E})$ [**H1**, 7.12, p. 162], we may take for the bundle N above the dual $\sigma^*\mathcal{O}(-1)$.

Remark. For the ruled surface $\mathbf{P}(E)$ over an elliptic curve C with $\deg\Lambda^2(E) = 1$, there are precisely three sub-line bundles of $S^2(E)$ of the correct degree: The $\mathcal{O}(p_i)$ in the proof of Theorem 2.

Remark. It is known [**H2**, 10.5, pp. 53-54] that, on a curve C of genus $g(C) \geq 2$, there are bundles of any given degree and rank that are stable and all of whose symmetric powers are also stable. For such a rank 2 bundle E, $E \otimes L$ has the same property for any line bundle L. Thus, even with E normalized if necessary we will have, for any sub-line bundle N of $S^n(E)$, $\deg(N) < (n/2)\deg \Lambda^2(E)$. For such a bundle E whose normalizations have positive degree, the monoid of effective divisor classes on the ruled surface $\mathbf{P}(E)$ requires an infinite number of generators.

References

[A] M.F. Atiyah, *Vector bundles over an elliptic curve*, Oxford University Press, Proceedings London Math. Soc., **7**(**3**) (1957), 414-452, MR 24 #A1274, Zbl 084.17305.

[H1] R. Hartshorne, *Algebraic Geometry*, Graduate Text in Mathematics, **52**, Springer-Verlag, 1977, MR 57 #3116, Zbl 367.14001.

[H2] ———, *Ample Subvarieties of Algebraic Varieties*, Lectures Notes in Math., **156**, Springer-Verlag, 1970, MR 44 #211, Zbl 208.48901.

[N] M. Nagata, *On self-intersection number of a section on a ruled surface*, Nagoya Math., **37** (1970), 191-196, MR 41 #3475 Zbl 193.21603.

[R1] J. Rosoff, *Effective divisor classes and blowings-up of* $\mathbf{P}^2$, Pacific J. Math., **89**(**2**) (1980), 419-429, MR 82e:14042, Zbl 564.14002.

[R2] ______, *The monoid of effective divisor classes on a complex torus*, Springer-Verlag, Lectures Notes in Math., **862** (1981), 208-231, MR 83m:14032, Zbl 488.14002.

Received August 12, 1999 and revised August 7, 2000. The author would like to thank the mathematics departments at the New Zeland Universities of Canterbury and Otago for their hospitality during the preparation of this manuscript, and to the referee for many valuable suggestions.

Gustavus Adolphus College
St. Peter, MN 56082
E-mail address: jr@gac.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 202, No. 1, 2002

ANISOTROPIC GROUPS OF TYPE A_n AND THE COMMUTING GRAPH OF FINITE SIMPLE GROUPS

Yoav Segev and Gary M. Seitz

In this paper we make a contribution to the Margulis-Platonov conjecture, which describes the normal subgroup structure of algebraic groups over number fields. We establish the conjecture for inner forms of anisotropic groups of type A_n. We obtain information on the commuting graph of nonabelian finite simple groups, and consequently, using the paper by Segev, 1999, we obtain results on the normal structure and quotient groups of the multiplicative group of a division algebra.

0. Introduction.

Let $\mathfrak{G}$ be a simple, simply connected algebraic group defined over an algebraic number field K. Let T be the (finite) set of all nonarchimedean places v of K such that $\mathfrak{G}$ is K_v-anisotropic, and define $\mathfrak{G}(K,T)$ to be $\prod_{v\in T}\mathfrak{G}(K_v)$ with the topology of the direct product if $T \neq \emptyset$, and let $\mathfrak{G}(K,T) = \{e\}$ if $T = \emptyset$ (which is always the case if $\mathfrak{G}$ is not of type A_n). Let $\delta : \mathfrak{G}(K) \to \mathfrak{G}(K,T)$ be the diagonal embedding in the first case, and the trivial homomorphism in the second case.

Conjecture (Margulis and Platonov)**.** *For any noncentral normal subgroup $N \leq \mathfrak{G}(K)$ there exists an open normal subgroup $W \leq \mathfrak{G}(K,T)$ such that $N = \delta^{-1}(W)$; in particular, if $T = \emptyset$, the group $\mathfrak{G}(K)$ has no proper noncentral normal subgroups* (*i.e., it is projectively simple*)*.*

The conjecture has been established for almost all isotropic groups and for most anisotropic groups except for those of type A_n. The anisotropic groups of type A_n are thus the main unresolved case of the conjecture.

Inner forms of anisotropic groups of type A_n have the form $SL_{1,D}$, the reduced norm 1 group of a finite dimensional division algebra D over K (see 2.17 and 2.12 of [**10**]). In this case Potapchik and Rapinchuk showed (Theorem 2.1 of [**11**]) that if $SL_{1,D}$ fails to satisfy the Conjecture, then there exists a proper normal subgroup N of $D^* = D - \{0\}$ such that D^*/N is a nonabelian finite simple group.

In recent work the first named author ([**14**]) established a result, relating finite simple images of the multiplicative group of a finite dimensional division algebra over an arbitrary field to information about the commuting graph of finite simple groups. To state this result we need the following definitions.

Let H be a finite group. The *commuting graph of H* denoted $\Delta(H)$ is the graph whose vertex set is $H - Z(H)$ and whose edges are pairs $\{h, g\} \subseteq H - Z(H)$, such that $h \neq g$ and $[h, g] \in Z(H)$. We denote the diameter of $\Delta(H)$ by $\mathrm{diam}\,(\Delta(H))$.

Let $d : \Delta(H) \times \Delta(H) \to \mathbb{Z}^{\geq 0}$ be the distance function on $\Delta(H)$. We say that $\Delta(H)$ is *balanced* if there exists $x, y \in \Delta(H)$ such that the distances $d(x, y)$, $d(x, xy)$, $d(y, xy)$, $d(x, x^{-1}y)$, $d(y, x^{-1}y)$ are all larger than 3.

Theorem (Segev [**14**]). *Let D be a finite dimensional division algebra over an arbitrary field and L a nonabelian finite simple group. If* $\mathrm{diam}\,(\Delta(L)) > 4$, *or $\Delta(L)$ is balanced, then L cannot be isomorphic to a quotient of D^*.*

Consequently, the Margulis-Platonov Conjecture for inner forms of anisotropic groups of type A_n is resolved by the following theorem, which is the main result of this paper.

Theorem 1. *Let L be a nonabelian finite simple group. Then either* $\mathrm{diam}\,(\Delta(L)) > 4$ *or $\Delta(L)$ is balanced.*

The following results are then immediate corollaries:

Theorem 2. *The Margulis-Platonov Conjecture holds for $\mathfrak{G} = \mathbf{SL}_{1,D}$.*

Theorem 3. *If D is a finite dimensional division algebra over an arbitrary field, then no quotient of D^* is a nonabelian finite simple group.*

In Section 12 we show that the following theorem is a consequence of Theorem 2.

Theorem 4. *Let D be a finite dimensional division algebra over a number field. Let N be a noncentral normal subgroup of D^*. Then D^*/N is a solvable group.*

To prove Theorem 1 we need to establish results on the commuting graph of a finite simple group. These results may have independent interest, so we state them as separate theorems corresponding to the various types of finite simple groups.

The main obstacle in establishing Theorem 1 occurs for classical groups. Here we prove the following theorem.

Theorem 5. *Let L be a finite simple group of classical type. Then $\Delta(L)$ is balanced. The required elements can be taken as opposite regular unipotent elements.*

Corollary. *If L is a finite simple classical group, then* $\operatorname{diam}(\Delta(L)) \geq 4$.

We mention that except for some small cases the elements x, y used to establish balance in Theorem 5 satisfy $d(x,y) = 4$ (see Section 12).

The following result covers exceptional groups of Lie type and Sporadic groups.

Theorem 6. *Let $L \not\simeq E_7(q)$ be either an exceptional group of Lie type or a Sporadic group. Then $\Delta(L)$ is disconnected. If $L = E_7(q)$, then $\Delta(L)$ is balanced, where the elements x, y can be chosen to be semisimple elements.*

For the alternating groups we have:

Theorem 7. *If L is a simple alternating group, then* $\operatorname{diam}(\Delta(L)) > 4$.

Finally, in Section 12 we prove the following theorem:

Theorem 8. *Let $G(q)$ be a simple classical group with $q > 5$. Then $\Delta(G(q))$ is disconnected if and only if one of the following holds*

(i) $G(q) \simeq L_n^\epsilon(q)$ *and n is a prime.*
(ii) $G(q) \simeq L_n^\epsilon(q)$, *$n-1$ is a prime and $q - \epsilon \mid n$.*
(iii) $G(q) \simeq S_{2n}(q), O_{2n}^-(q)$, *or $O_{2n+1}(q)$ and $n = 2^c$, for some c.*

Moreover, if $\Delta(G(q))$ is connected then $\operatorname{diam}(\Delta(G(q))) \leq 10$.

We draw the attention of the reader to the remark at the end of Section 12, for additional information about the connectivity of the commuting graph of finite simple groups.

In Chapter 1, which consists of Sections 1-7 we prove Theorem 5. In Chapter 2, which consists of Sections 8-9 we prove Theorem 6, when L is an exceptional group of Lie-type. Section 10 is devoted to the Alternating groups and the short Section 11 is devoted to the Sporadic groups. Finally in Section 12 we derive Theorem 4 from Theorem 2 and we include some results and remarks about the commuting graph of the classical groups.

We would like to thank Michael Aschbacher for various discussions, in particular, for contributions in Sections 8 and 9.

Chapter 1. The Classical Groups.

1. Notation and preliminaries.

The notation and definitions that will be introduced in this section will prevail *throughout Chapter* 1. $\mathbb{F}$ denotes a finite field and V denotes a vector space of dimension n over $\mathbb{F}$. We fix an ordered basis

$$\mathcal{B} = \{v_1, \ldots, v_n\}$$

of V. For a subset $S \subseteq V$, $\langle S \rangle$ denotes the subspace generated by S. We set:

$$\text{For } 1 \le i \le n, \quad \mathcal{V}_i = \langle v_1, v_2, \dots, v_i \rangle.$$

We write $M(V)$ for both $\mathrm{Hom}_{\mathbb{F}}(V, V)$, the set of all linear operators on V, and for the set of $n \times n$ matrices over $\mathbb{F}$. When we wish to emphasize that we are dealing with matrices we'll write $M_n(\mathbb{F})$ for the set of $n \times n$ matrices over $\mathbb{F}$. Also $GL(V) \subseteq M(V)$, denotes both the set of invertible linear operators on V and the set of invertible $n \times n$ matrices over $\mathbb{F}$. To emphasize matrices we write $GL_n(\mathbb{F})$, for the set of $n \times n$ invertible matrices over $\mathbb{F}$. Finally, $SL(V) \subseteq M(V)$ are the elements of determinant 1; again, we write $SL_n(\mathbb{F})$ for the set of $n \times n$ matrices of determinant 1. We use the same notation for the linear operator and its matrix, *with respect to the basis $\mathcal{B}$*. All our matrices are also linear operators whose matrix is the given matrix always with respect to our fixed basis $\mathcal{B}$, unless explicitly mentioned otherwise. Thus if $a \in M(V)$, then a is an $n \times n$ matrix over $\mathbb{F}$ whose (i, j)-th entry we always denote by a_{ij}. Also $a : V \to V$ is a linear operator such that $v_i a = \sum_{j=1}^{n} a_{ij} v_j$.

Given a bilinear form f (resp. a quadratic form Q) on V, we denote by $O(V, f)$ (resp. $O(V, Q)$) the elements in $GL(V)$ preserving f (resp. Q). $SO(V, f)$ (resp. $SO(V, Q)$) denotes the elements in $O(V, f)$ (resp. $O(V, Q)$) of determinant 1.

We fix *the letter* $\mathcal{R}$ to denote either $\mathbb{F}$, or the ring of polynomials over $\mathbb{F}, \mathbb{F}[\lambda]$. We'll denote by $M_n(\mathcal{R})$, the set of $n \times n$ matrices over $\mathcal{R}$.

Let H be a finite group. The *commuting graph of H* denoted $\Delta(H)$ is the graph whose vertex set is $H - Z(H)$ and whose edges are pairs $\{h, g\} \subseteq H - Z(H)$, such that $h \neq g$ and $[h, g] \in Z(H)$. (*Note* that our definition of the commuting graph differs a bit from what the reader may be used to, i.e., the vertex set of $\Delta(H)$ is $H - Z(H)$ and not $H - \{1\}$ and two elements form an edge when they commute *modulo the center of H* and not only when they commute.) We denote by $d_{\Delta(H)}$ the distance function of $\Delta(H)$. We fix the letter Δ to denote $\Delta(GL(V))$ and the letter d to denote the distance function of Δ (see 1.3 for further notation and definitions for the commuting graph).

Our goal in Chapter 1 is to prove Theorem 5 of the Introduction, which shows that $\Delta(L)$ is balanced, for all simple classical groups L. In principle we present a uniform approach to this, by showing that in all cases we can take the elements x, y to be opposite regular unipotent elements. However, the details are fairly complicated. In this section and the next we lay the ground work for the proof.

1.1. Notation and definitions for matrices over $\mathcal{R}$. Let $m \geq 1$ be an integer.

(1) First we mention that given $\alpha \in \mathbb{F}$, whenever we write $\overline{\alpha}$ inside a matrix, this means $\overline{\alpha} = -\alpha$.
(2) I_m denotes the identity $m \times m$ matrix.
(3) For integers $i, j \geq 1$, $0_{i,j}$ denotes the zero $i \times j$ matrix. We denote by 0_i the zero $i \times i$ matrix.
(4) Given $g \in M_m(\mathbb{F})$, we denote the transpose of g by g^t.
(5) Given $A \in M_m(\mathcal{R})$, $M_{i,j}(A) \in M_{m-1}(\mathcal{R})$, denotes the (i,j)-minor of A, i.e., the matrix A without the i-th row and j-th column. Also $M_{(i_1,i_2),(j_1,j_2)}(A) \in M_{m-2}(\mathcal{R})$ is the matrix without the i_1, i_2 rows and without the j_1, j_2 columns.
(6) Suppose $m = k_1 + k_2 + \cdots + k_t$ and that $g_i \in M_{k_i}(\mathcal{R})$, $1 \leq i \leq t$. We write $g = \operatorname{diag}(g_1, g_2, \ldots, g_t)$ for the $m \times m$ matrix with $g_1, g_2, \ldots, g_t$ on the main diagonal (in that order) and zero elsewhere. Of course if $g_i \in \mathcal{R}$, for all i ($k_i = 1$, for all i), then g is a diagonal matrix in the usual sense.
(7) Suppose $m \geq 2$ and let $1 \leq i \leq m-1$ and $\alpha \in \mathbb{F}$. We denote by $u_i^m(\alpha) \in M_m(\mathbb{F})$, the matrix which has 1 on the main diagonal, α in the $(i+1, i)$ entry and zero elsewhere.
(8) Suppose $m \geq 2$ and let $\beta_1, \beta_2, \ldots, \beta_{m-1} \in \mathbb{F}^*$. We denote

$$a_m(\beta_1, \beta_2, \ldots, \beta_{m-1}) = u_1^m(\beta_{m-1}) u_2^m(\beta_{m-2}) \cdots u_{m-2}^m(\beta_2) u_{m-1}^m(\beta_1)$$
$$b_m(\beta_1, \beta_2, \ldots, \beta_{m-1}) = u_1^m(-\beta_1) u_2^m(-\beta_2) \cdots u_{m-2}^m(-\beta_{m-2}) u_{m-1}^m(-\beta_{m-1}).$$

Of course

$$a_m(\beta_1, \ldots, \beta_{m-1}) = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \beta_{m-1} & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \beta_{m-2} & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \beta_{m-3} & 1 & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & \beta_2 & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \beta_1 & 1 \end{bmatrix},$$

$$b_m(\beta_1, \ldots, \beta_{m-1}) = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \overline{\beta}_1 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \overline{\beta}_2 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \overline{\beta}_3 & 1 & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & \overline{\beta}_{m-2} & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \overline{\beta}_{m-1} & 1 \end{bmatrix}.$$

(9) We denote $a_1 = b_1 = [1]$ and for $m \geq 2$,

$$a_m = a_m(1, 1, \ldots, 1) \quad \text{and} \quad b_m = b_m(1, 1, \ldots, 1).$$

Hence

$$a_m = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & 1 & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & & 0 & 1 & 1 \end{bmatrix} \qquad b_m = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \bar{1} & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \bar{1} & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \bar{1} & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & \bar{1} & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & & 0 & \bar{1} & 1 \end{bmatrix}.$$

(10) Suppose $m \geq 2$ and $1 \leq r \leq m-1$. We denote by $\mathcal{T}_m(r)$ the set of $m \times m$ matrices $t \in M_m(\mathbb{F})$ such that:
(i) $t_{i,j} = 0$, for all $1 \leq i \leq r$ and $1 \leq j \leq m$.
(ii) $t_{r+i,i} \neq 0$ and $t_{r+i,\ell} = 0$, for all $1 \leq i \leq m-r$ and all $i < \ell \leq m$.
Thus t has the form

$$t = \begin{bmatrix} 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ t_{r+1,1} & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ * & t_{r+2,2} & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ * & * & t_{r+3,3} & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ * & * & * & * & * & t_{m,m-r} & 0 & \cdot & \cdot & 0 \end{bmatrix}$$

where $*$ represents any element of $\mathbb{F}$.

(11) Throughout Chapter 1, J_m denotes the following $m \times m$ matrix. If we set, $J = J_m$, then $J_{i,m+1-i} = (-1)^{i+1}$, for all $1 \leq i \leq m$, and $J_{i,j} = 0$, otherwise. Thus

$$J_m = \begin{bmatrix} 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & \bar{1} & 0 \\ 0 & 0 & \cdot & \cdot & 0 & 1 & 0 & 0 \\ 0 & 0 & \cdot & 0 & \bar{1} & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \bar{1}^m & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \bar{1}^{m+1} & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix}.$$

Note that $J_m^{-1} = J_m^t$, $J_m^2 = (-1)^{m+1} I_m$ and if $m = 2\ell$ is even, then

$$J_{2\ell} = \begin{bmatrix} 0_\ell & J_\ell \\ (-1)^\ell J_\ell & 0_\ell \end{bmatrix}.$$

1.2. Notation for polynomials, characteristic polynomials and characteristic vectors. Let $m \geq 1$ be an integer.

(1) Let $g \in M_m(\mathbb{F})$. We denote by $F_g[\lambda]$, the characteristic polynomial of g. We often write F_g for $F_g[\lambda]$.
(2) If F is the characteristic polynomial of $g \in GL_m(\mathbb{F})$, we denote by $\overline{F}$ the characteristic polynomial of g^{-1}.
(3) Given a polynomial $F[\lambda]$, we denote by $\alpha(F, \ell)$, the coefficient of λ^ℓ in F.
(4) Throughout Chapter 1 we denote by $F_m[\lambda]$ the characteristic polynomial of $a_m^t a_m$ (a_m as in 1.1.9). We mention that several properties of $F_m[\lambda]$ are given in 2.6.
(5) Throughout Chapter 1, $G_m[\lambda]$ denotes the characteristic polynomial of the following $m \times m$ matrix

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

(6) We denote $Q_m[\lambda] = \lambda^m - \lambda^{m-1} + \lambda^{m-2} + \cdots + (-1)^{m-1}\lambda + (-1)^m$.
(7) Let $g \in GL(V)$ and suppose that $v \in V$ is a characteristic vector for g. We denote by $\lambda_g(v) \in \mathbb{F}$ the scalar such that $vg = \lambda_g(v)v$.

1.3. Notation for the commuting graph. Let H be a group and let $\Lambda = \Delta(H)$.

(1) Given elements $X, Y \in \Lambda$, we write $B_\Lambda(X, Y)$ if the distances $d_\Lambda(X, Y)$, $d_\Lambda(X, XY)$ and $d_\Lambda(X, X^{-1}Y)$ are all > 3. We write $B(X, Y) = B_\Delta(X, Y)$ (recall that $\Delta = \Delta(GL(V))$).
(2) We say that Λ is *balanced* if there are elements $X, Y \in \Lambda$ such that $B_\Lambda(X, Y)$ and $B_\Lambda(Y, X)$.
(3) We use the usual notation for graphs, thus, for example, $\Delta^{\leq i}(X)$ means the set of all elements at distance at most i from X, in Δ.

1.4. Further notation and definitions. Let $g \in GL(V)$, $0 \neq v \in V$ and $H \leq GL(V)$, a subgroup.

(1) We denote by $\mathcal{O}(v, g)$ the orbit of v under $\langle g \rangle$.
(2) Given an ordered basis $\mathcal{A} = \{w_1, \ldots, w_n\}$ of V we denote by $[g]_\mathcal{A}$ the matrix of g with respect to the basis $\mathcal{A}$. Thus, the i-th *row* of $[g]_\mathcal{A}$ are the coordinates of $w_i g$ with respect to $\mathcal{A}$.
(3) We say that H is *closed under transpose* if $h \in H$ implies $h^t \in H$.

(4) We fix the letter τ to denote the graph automorphism of $SL_n(\mathbb{F})$ such that $\tau : u_i^n(\alpha) \to u_{n-i}^n(\alpha)$ and $\tau : (u_i^n(\alpha))^t \to (u_{n-i}^n(\alpha))^t$, for all $\alpha \in \mathbb{F}$ and all $1 \le i \le n-1$. Note that τ commutes with the transpose map.
(5) If $|\mathbb{F}| = q^2$, we let $\sigma_q : GL_n(\mathbb{F}) \to GL_n(\mathbb{F})$, be the Frobenius automorphism taking each entry of $g \in GL_n(\mathbb{F})$ to its q power.

By a *Classical Group* we mean $L \le GL(V)$, where L is one of the groups $SL_n(q)$, $Sp_n(q)$, $\Omega_n^\epsilon(q)$, or $SU_n(q)$, where for orthogonal groups we use $\epsilon \in \{+,-\}$ only in even dimension and for unitary groups we work over the field of order q^2. In all cases we take L to be quasisimple, avoiding the few cases when this does not hold. By a *Simple Classical Group* we mean $L/Z(L)$, with L a classical group. In the respective cases we denote the simple classical groups by $L_n(q), S_n(q), O_n(q), O_n^\epsilon(q)$ and $U_n(q)$.

1.5. (1) *For even q and odd n, $O_n(q) \simeq S_{n-1}(q)$.*
(2) *For all q, $O_3(q) \simeq L_2(q)$, $O_4^+(q) \simeq L_2(q) \times L_2(q)$, $O_4^-(q) \simeq L_2(q^2)$, $O_5(q) \simeq S_4(q)$, $O_6^+(q) \simeq L_4(q)$ and $O_6^-(q) \simeq U_4(q)$.*

The purpose of Chapter 1 is to prove:

Theorem 1.6. *Let L be a finite simple classical group. Then $\Delta(L)$ is balanced.*

We mention that in Remark 1.18 ahead we indicate our strategy for proving Theorem 1.6.

1.7. *Let H be a group. Suppose that $Z(H/Z(H)) = 1$ and that $\Delta(H)$ is balanced. Then $\Delta(H/Z(H))$ is balanced.*

Proof. This is obvious since if $X, Y \in \Delta(H)$ satisfy $B(X,Y)$ and $B(Y,X)$, then $XZ(H)$, $YZ(H)$ satisfy the same condition in $\Delta(H/Z(H))$.

1.8. *Let $L \le SL(V)$ be a classical group. Set $\Lambda = \Delta(L)$ and suppose that L is closed under transpose. Then:*
(1) *The maps $g \to g^{-1}$, $g \to g^t$ and conjugation are isomorphisms of Λ.*
(2) *Let $g, h \in \Lambda$ and let $\epsilon \in \{1,-1\}$, then any one of the following imply $d_\Lambda(g, g^\epsilon h) > 3$:*
(i) $d_\Lambda(g, hg^\epsilon) > 3$;
(ii) $d_\Lambda(g, h^{-1}g^{-\epsilon}) > 3$;
(iii) $d_\Lambda(g, g^{-\epsilon}h^{-1}) > 3$.

Proof. (1) is easy. (2) follows from (1) noting that $(g^\epsilon h)^{g^\epsilon} = hg^\epsilon$, $(g^{-\epsilon}h^{-1})^{g^{-\epsilon}} = h^{-1}g^{-\epsilon}$ and that the distance between g and t is the same as that from g to t^{-1}.

1.9. *Let $L \le SL(V)$ be a classical group. Set $\Lambda = \Delta(L)$ and suppose that L is closed under transpose. Let $X, Y \in L$. Then:*

(1) *If* $B(X,Y)$, *then* $B(X^t,Y^t)$.
In particular:
(2) *If* $B(X,X^t)$, *then* $B(X^t,X)$.

Proof. Suppose that $B(X,Y)$ holds. By 1.8.1, $d_\Lambda(X^t,Y^t) > 3$. Also since $d_\Lambda(X,XY) > 3$, $d_\Lambda(X^t,(XY)^t) > 3$. Hence $d_\Lambda(X^t,Y^tX^t) > 3$. By 1.8.2, $d_\Lambda(X^t,X^tY^t) > 3$. Finally since $d_\Lambda(X,X^{-1}Y) > 3$, $d_\Lambda(X^t,(X^{-1}Y)^t) > 3$. Thus $d_\Lambda(X^t,Y^t(X^t)^{-1}) > 3$ and then, $d_\Lambda(X^t,(X^t)^{-1}Y^t) > 3$.

Corollary 1.10. *Let $L \leq SL(V)$ be a classical group. Set $\Lambda = \Delta(L)$ and suppose that L is closed under transpose. Suppose one of the following holds:*
(i) *There exists $X \in L$ such that $B_\Lambda(X,X^t)$.*
(ii) *There exists $X,Y \in L$ such that $B_\Lambda(X,Y^t)$ and $B_\Lambda(Y,X^t)$.*
Then $\Delta(L)$ is balanced.

Proof. If (i) holds, then it is immediate from 1.9.2, and definition, that $\Delta(L)$ is balanced. If (ii) holds, then by 1.9.1, also $B_\Lambda(Y^t,X)$, so by definition $\Delta(L)$ is balanced.

1.11. *Suppose $n = 2k + \epsilon \geq 2$, with $\epsilon \in \{0,1\}$. Let $\beta_1,\beta_2,\dots,\beta_{k-1} \in \mathbb{F}^*$. Set $a = a_k(\beta_1,\beta_2,\dots,\beta_{k-1})$ and $b = b_k(\beta_1,\beta_2,\dots,\beta_{k-1})$. Let $\tau : SL_n(\mathbb{F}) \to SL_n(\mathbb{F})$ be the automorphism defined in 1.4.4. If $\epsilon = 0$, then* $\mathrm{diag}\,(a,b^{-1}) \in \mathrm{Fix}\,(\tau)$ *and if $\epsilon = 1$, then* $\mathrm{diag}\,(a,1,b^{-1}) \in \mathrm{Fix}\,(\tau)$.

Proof. Just observe that if $\epsilon = 0$, then

$$\begin{aligned}&\mathrm{diag}\,(a,b^{-1})\\&= u_1^n(\beta_{k-1})u_{n-1}^n(\beta_{k-1})u_2^n(\beta_{k-2})u_{n-2}^n(\beta_{k-2})\cdots u_{k-1}^n(\beta_1)u_{k+1}^n(\beta_1)\end{aligned}$$

and if $\epsilon = 1$, then

$$\begin{aligned}&\mathrm{diag}\,(a,1,b^{-1})\\&= u_1^n(\beta_{k-1})u_{n-1}^n(\beta_{k-1})u_2^n(\beta_{k-2})u_{n-2}^n(\beta_{k-2})\cdots u_{k-1}^n(\beta_1)u_{k+2}^n(\beta_1).\end{aligned}$$

1.12. *Let $\tau,\sigma_q : SL(V) \to SL(V)$ be the automorphisms defined in 1.4.4 and 1.4.5. Set $J = J_n$ (see* 1.1.11*). Then:*
(1) *$g\tau = J(g^t)^{-1}J^{-1} = J(g^t)^{-1}J^t$, for all $g \in SL(V)$.*
(2) *τ and σ_q commute with the transpose map.*
(3) *For an automorphism $\phi : SL(V) \to SL(V)$, let* $\mathrm{Fix}\,(\phi) = \{h \in SL(V) : h\phi = h\}$. *Then if $|\mathbb{F}| = q^2$,* $\mathrm{Fix}\,(\tau\sigma_q) \simeq SU_n(q)$*; if n is even, then* $\mathrm{Fix}\,(\tau) \simeq Sp_n(q)$ *and if n is odd and q is odd,* $\mathrm{Fix}\,(\tau) \simeq SO_n(q)$.
(4) *In the notation of (3),* $\mathrm{Fix}\,(\tau)$ *and* $\mathrm{Fix}\,(\tau\sigma_q)$ *are closed under transpose.*
(5) *Suppose $n = 2k$ is even, $x,y \in SL_k(\mathbb{F})$ are such that* $\mathrm{diag}\,(x,y^{-1}) \in \mathrm{Fix}\,(\tau)$. *Then $y = J_kx^tJ_k^{-1} = J_kx^tJ_k^t$.*

Proof. First recall that $J^{-1} = J^t$. Let $\tau' : SL(V) \to SL(V)$, be the automorphism $g \to J(g^t)^{-1}J^{-1}$. It is easy to check that $u_i^n(\alpha)\tau' = u_i^n(\alpha)\tau$, and $(u_i^n(\alpha))^t\tau' = (u_i^n(\alpha))^t\tau$, for all $1 \leq i \leq n-1$, and all $\alpha \in \mathbb{F}$. Thus $\tau' = \tau$.

Evidently τ and σ_q commute with the transpose map. Next note that $g \in \mathrm{Fix}\,(\tau)$ iff $gJg^t = J$; thus $g \in SO(V,f)$, where f is the bilinear form given by $f(v_i, v_j) = J_{i,j}$. Hence $\mathrm{Fix}\,(\tau)$ is as claimed in (3). Now if $|\mathbb{F}| = q^2$, then $g \in \mathrm{Fix}\,(\tau\sigma_q)$ iff $gJ(g\sigma_q)^t = J$, so as above, $g \in SO(V,f)$, for a suitable unitary form f.

To prove (5), set $g = \mathrm{diag}\,(x, y^{-1})$. Then by (1), $g\tau = J(g^t)^{-1}J^t = J\mathrm{diag}\,((x^t)^{-1}, y^t)J^t$. Now using Definition 1.1.11, we get

$$\begin{aligned} g\tau &= \begin{bmatrix} 0_k & J_k \\ (-1)^k J_k & 0_k \end{bmatrix} \cdot \begin{bmatrix} (x^t)^{-1} & 0_k \\ 0_k & y^t \end{bmatrix} \cdot \begin{bmatrix} 0_k & (-1)^k J_k^t \\ J_k^t & 0_k \end{bmatrix} \\ &= \begin{bmatrix} 0_k & J_k y^t \\ (-1)^k J_k (x^t)^{-1} & 0_k \end{bmatrix} \cdot \begin{bmatrix} 0_k & -J_k \\ (-1)^{k+1} J_k & 0_k \end{bmatrix} \\ &= \begin{bmatrix} (-1)^{k+1} J_k y^t J_k & 0_k \\ 0_k & (-1)^{k+1} J_k (x^t)^{-1} J_k \end{bmatrix}. \end{aligned}$$

Since we are assuming that $g\tau = g$, we see that $(-1)^{k+1}J_k y^t J_k = x$, so since $J_k^{-1} = (-1)^{k+1}J_k = J_k^t$, we see that $x = J_k^t y^t J_k$, so $y = J_k x^t J_k^{-1} = J_k x^t J_k^t$, as asserted.

1.13. *Let $X \in GL_n(V)$ be a lower triangular matrix such that $X - I_n \in \mathcal{T}_n(1)$ (see 1.1.10 for $\mathcal{T}_n(1)$). Let $h \in M_n(\mathbb{F})$ be a matrix commuting with X. Then:*

(1) *h is a lower triangular matrix.*
(2) *There exists $1 \le r < n$, and $\beta \in \mathbb{F}$ such that $h - \beta I_n \in \mathcal{T}_n(r)$.*
(3) *If $X_{i,i-1} = X_{j,j-1}$, for all $2 \le i,\ j \le n$, then $h_{r+i,i} = h_{r+j,j}$, for all $1 \le i, j \le n - r$.*

Proof. For $2 \le i \le n$, set $\alpha_i := X_{i,i-1}$. Note that by definition (see 1.1.10), $\alpha_i \neq 0$, for all $2 \le i \le n$. Note further that h commutes with the matrix $X - I_n$, and clearly for $1 \le i \le n-1$, $\ker(X - I_n)^i = \mathcal{V}_i$. Since h commutes with $(X - I_n)^i$, h fixes $\ker(X - I_n)^i$, so (1) holds.

Next set $Xh = g$ and $hX = q$. It is easy to check that for $2 \le i \le n$, $g_{i,i-1} = \alpha_i h_{i-1,i-1} + h_{i,i-1}$ and that $q_{i,i-1} = h_{i,i-1} + \alpha_i h_{i,i}$. Since $g = q$, and $\alpha_i \neq 0$, for all i, we see that $h_{1,1} = h_{2,2} = \cdots = h_{n,n}$. Set $\beta = h_{1,1}$ and $t = h - \beta I_n$. Then t has the form

$$t = \begin{bmatrix} 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ t_{r+1,1} & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ * & t_{r+2,2} & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ * & * & t_{r+3,3} & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ * & * & * & * & * & t_{n,n-r} & 0 & \cdot & \cdot & 0 \end{bmatrix}$$

where $1 \le r \le n-1$ and for some $1 \le j \le n-r$, $t_{r+j,j} \ne 0$. Note that $X - I_n$ commutes with t.

Set $(X-I_n)t = g$ and $t(X-I_n) = q$. Then it is easy to check that $g_{r+2,1} = \alpha_{r+2}t_{r+1,1}$, $g_{r+3,2} = \alpha_{r+3}t_{r+2,2}, \dots, g_{n,n-r-1} = \alpha_n t_{n-1,n-r-1}$. Similarly, $q_{r+2,1} = \alpha_2 t_{r+2,2}$, $q_{r+3,2} = \alpha_3 t_{r+3,3}, \dots, q_{n,n-r-1} = \alpha_{n-r}t_{n,n-r}$. Since $g = q$, $\alpha_i \ne 0$, for all i, and $t_{r+j,j} \ne 0$, for some $1 \le j \le n-r$, $t_{r+i,i} \ne 0$, for all $1 \le i \le n-r$ and $t \in \mathcal{T}_n(r)$ as asserted. Further, it is easy to check that (3) holds.

1.14. *Let $R, S \in GL(V)$. Set $\mathfrak{Z} = Z(GL(V))$ and $\mathcal{W} = \langle \mathcal{O}(w_1, S)\rangle$. Suppose that:*

(a) *$R^{-1}SR = \mu S$, for some $\mu \in \mathbb{F}^*$.*
(b) *v_1 is a characteristic vector of R.*

Then:

(1) *If $\mu = 1$, then $\mathcal{W}$ is a set of characteristic vectors of R and for $w \in \mathcal{W}$, $\lambda_R(w) = \lambda_R(v_1)$. In particular, if $\mathcal{W} = V$, then $R \in \mathfrak{Z}$.*

Suppose $\mathcal{W} = V$, and let $F_S[\lambda] = \lambda^n - \sum_{i=0}^{n-1} \alpha_i\lambda^i$. Then:

(2) *R is conjugate in $GL(V)$ to some member of* $\operatorname{diag}(1, \mu, \mu^2, \dots, \mu^{n-1})\mathfrak{Z}$.
(3) *$\mu^i = 1$, for each $1 \le i \le n$ such that $\alpha_{n-i} \ne 0$.*
(4) *$\mu^n = 1$.*
(5) *If gcd $\Big\{\{i : \alpha_{n-i} \ne 0\} \cup \{|\mathbb{F}^*|\}\Big\} = 1$, then $R \in \mathfrak{Z}$.*

Proof. Notice that by hypotheses (a) and (b), $\mathcal{O}(v_1, S)$ is a set of characteristic vectors of R. Further if $\mu = 1$, clearly (1) holds. For the remaining parts assume $\mathcal{W} = V$. Then $\mathcal{A} = \{v_1, v_1S, v_1S^2, \dots, v_1S^{n-1}\}$ is a basis of V. The matrix of S with respect to the basis $\mathcal{A}$ is

$$S' := [S]_{\mathcal{A}} = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 1 \\ \alpha_0 & \alpha_1 & \alpha_2 & \cdot & \cdot & \cdot & \cdot & \alpha_{n-1} \end{bmatrix}$$

and the matrix of R with respect to the basis $\mathcal{A}$ is $R' = \operatorname{diag}(R_1, R_2, \dots, R_n)$. Replacing R with a scalar multiple of R we may assume that $R_1 = 1$. Note that for $1 \le i \le n-1$, the $(i, i+1)$-entry of the matrix $(R')^{-1}S'R'$ is $R_i^{-1}R_{i+1}$. Since $(R')^{-1}S'R' = \mu S'$, we conclude that $R_i = \mu^{i-1}$, $1 \le i \le n$ and (2) holds.

Next note that for $1 \le i \le n$, the $(n, n-i+1)$-entry of $(R')^{-1}S'R'$ is $R_n^{-1}R_{n-i+1}\alpha_{n-i} = \mu^{1-n}\mu^{n-i}\alpha_{n-i} = \mu^{1-i}\alpha_{n-i}$. Thus, since $(R')^{-1}S'R' = \mu S'$, $\mu^{1-i}\alpha_{n-i} = \mu\alpha_{n-i}$, so if $\alpha_{n-i} \ne 0$, $\mu^i = 1$. This shows (3). Of course

(4) follows from (3), since $\alpha_0 = (-1)^{n+1}\det(R) \neq 0$. Finally (5) is an immediate consequence of (2), (3) and (4).

1.15. *Suppose $S, T \in M(V)$, $R \in GL(V)$ and $j, m, \ell \geq 0$ are integers such that:*

(a) $1 \leq j \leq n-1$ *and for all* $1 \leq i \leq j$ *and* $i+1 < k \leq n$, $S_{i,i+1} \neq 0$ *and* $S_{i,k} = 0$.
(b) $\mathcal{V}_j \subseteq \ker(T)$.
(c) $v_{j+1} \notin \ker(S^\ell T)$.
(d) $1 \leq m \leq j+1$, *and $\mathcal{V}_m$ is R-invariant.*
(e) *If we set $\mathfrak{Z} = Z(GL(V))$ then $R^{-1}SR \in \mathfrak{Z}S$ and $R^{-1}TR \in \mathfrak{Z}T$.*

Then v_1 is a characteristic vector of R.

Proof. For $i \geq 0$, set $z_i = S^i T$. Note that $R^{-1}z_i R \in \mathfrak{Z}z_i$, for all $i \geq 0$ and hence

(i) $$\ker(z_i) \text{ is } R\text{-invariant, for all } i \geq 0.$$

Notice that by (a):

(ii) $$\text{For all } i \geq 0, \text{ if } \mathcal{V}_{j+1} \subseteq \ker(z_i), \text{ then } \mathcal{V}_j \subseteq \ker(z_{i+1}).$$

Now without loss we may assume that ℓ is the least nonnegative integer i such that $v_{j+1} \notin \ker(z_i)$. Since by (b), $\mathcal{V}_j \subseteq \ker(z_0)$, minimality of ℓ and (ii) imply that $\mathcal{V}_j \subseteq \ker(z_\ell)$. Thus

(iii) $$v_{j+1} \notin \ker(z_\ell) \quad \text{and} \quad \mathcal{V}_j \subseteq \ker(z_\ell).$$

Now, by (a) and (iii), we get that

(iv) $$\ker(z_{\ell+i}) \cap \mathcal{V}_{j-i+1} = \mathcal{V}_{j-i}, \qquad \text{for all } 0 \leq i \leq j-1.$$

By (i), (iv), (d) and since $1 \leq m \leq j+1$, we see that $\mathcal{V}_m, \mathcal{V}_{m-1}, \dots, \mathcal{V}_1$ are all R-invariant, so since $\mathcal{V}_1$ is R-invariant, v_1 is a characteristic vector of R.

1.16. *Suppose $n \geq 2$ and let $Z \in GL(n, \mathbb{F})$. Let $v \in V$ such that $\langle \mathcal{O}(v, Z) \rangle = V$ and let $\alpha \in \mathbb{F}$. Then $\langle \mathcal{O}(\alpha v + vZ, Z) \rangle \neq V$ iff $-\alpha$ is a characteristic value of Z.*

Proof. Since $\langle \mathcal{O}(v, Z) \rangle = V$, $\mathcal{C} := \{v, vZ, \dots, vZ^{n-1}\}$ is a basis of V. Now $\langle \mathcal{O}(\alpha v + vZ, Z) \rangle = V$, iff $\mathcal{D} := \{\alpha v + vZ, (\alpha v + vZ)Z, \dots, (\alpha v + vZ)Z^{n-1}\}$ is a basis of V. Now $\mathcal{D}$ is obtained from $\mathcal{C}$ by applying the transformation $\alpha I_n + Z$ to the basis $\mathcal{C}$. Thus $\mathcal{D}$ is a basis of V iff $\alpha I_n + Z$ is invertible and the lemma follows.

Corollary 1.17. *Suppose $n = 2k+1$ (with $k \geq 1$), let $S \in GL(n, \mathbb{F})$ and write*

$$S = \begin{bmatrix} R_{1,1} & R_{1,2} \\ R_{2,1} & Z \end{bmatrix}$$

with $R_{1,1}, R_{1,2}, R_{2,1}$ and Z a $k \times k$, $k \times (k+1)$, $(k+1) \times k$ and $(k+1) \times (k+1)$ matrices, respectively. Set $\mathcal{W} = \langle \mathcal{O}(v_1, S) \rangle$ and assume:

(a) $\mathcal{V}_k \subseteq \mathcal{W}$.
(b) $Z \in GL_{k+1}(\mathbb{F})$ *and* $\langle \mathcal{O}(v_{k+1}, \mathrm{diag}\,(I_k, Z))\rangle = \langle v_{k+1}, \dots, v_n\rangle$.
(c) $\alpha v_{k+1} + v_{k+1}\mathrm{diag}\,(I_k, Z) \in \mathcal{W}$*, for some* $\alpha \in \mathbb{F}$.

If $-\alpha$ *is not a characteristic value of the matrix* Z*, then* $V = \langle \mathcal{O}(v_1, S)\rangle$.

Proof. Set $\mathcal{U} = \langle v_{k+1}, \dots, v_n\rangle$ and let Z denote also the linear operator $Z : \mathcal{U} \to \mathcal{U}$, given by the matrix Z, with respect to the basis $\{v_{k+1}, \dots, v_n\}$. Then, by (b), $\mathcal{U} = \langle \mathcal{O}(v_{k+1}, Z)\rangle$. Also it is easy to check that hypothesis (a) implies that if $u \in \mathcal{U} \cap \mathcal{W}$, then $uZ \in \mathcal{U} \cap \mathcal{W}$. Hence by hypothesis (c), $\mathcal{O}(\alpha v_{k+1} + v_{k+1}Z, Z) \subseteq \mathcal{W}$. Now 1.16 and hypotheses (b) and (c) imply that if $-\alpha$ is not a characteristic value of Z, then $\mathcal{U} \subseteq \mathcal{W}$, so by (a), $\mathcal{W} = V$ as asserted.

1.18. Important remark. Throughout Chapter 1, the following strategy will be used to prove Theorem 1.6. Let $L \leq SL(V)$ be a classical group. Let $\Lambda = \Delta(L)$. We carefully choose $X, Y \in \Lambda$. To show $B_\Lambda(X, Y)$, let $S \in \{Y, XY, X^{-1}Y\}$. In order to show that $d_\Lambda(X, S) > 3$, suppose $R \in \Lambda^{\leq 2}(X) \cap \Lambda^{\leq 1}(S)$. We do the following steps.

Step 1. We obtain information about $C_L(X)$. Part of the work was already done in 1.13.

Step 2. Using Step 1, we show that if $h \in \Lambda^{\leq 1}(X) \cap \Lambda^{\leq 1}(R)$, then there exists $\beta \in \mathbb{F}^*$ and an integer $k \geq 1$ such that if we set $T := (h - \beta I_n)^k$, then there are integers $j, \ell, m \geq 0$ such that T, S, R, j, ℓ, m satisfy all the hypotheses of 1.15. Thus we conclude from 1.15 that v_1 is a characteristic vector of R.

Step 3. We compute $\langle \mathcal{O}(v_1, S)\rangle$. In all cases X, Y are chosen so that either $\langle \mathcal{O}(v_1, S)\rangle = V$, or $[S, R] = 1$, (so that we can use 1.14.1) and $\langle \mathcal{O}(v_1, S)\rangle$ has codimension 1 or 2 in V.

Step 4. We obtain information on the characteristic polynomial of S. This information is aimed to fit the hypotheses of 1.14.5.

Step 5. We use Step 2, Step 3 and Step 4, together with 1.14, to get that $R \in Z(L)$ and obtain a contradiction.

2. Some information about characteristic polynomials.

Throughout this section $n = 2k+\epsilon \geq 2$ is a positive integer, where $\epsilon \in \{0, 1\}$. a_m and b_m are as in 1.1.9. We draw the attention of the reader to 1.1 and 1.2, where we fixed our notation for matrices and polynomials. In particular, recall that the polynomials $F_m[\lambda], G_m[\lambda]$ and $Q_m[\lambda]$ are defined in 1.2.4, 1.2.5 and 1.2.6 respectively.

2.1. Notation. For an integer $\ell \geq 1$ and a prime r, $|\ell|_r$ is the largest power of r dividing ℓ. Hence, if $\gcd(\ell, r) = 1$, then $|\ell|_r = 0$.

2.2. *Let $\ell \geq 1$ be a positive integer. Suppose $\ell = \sum_{i=0}^{s} \epsilon_i 2^i$, with $\epsilon_i \in \{0,1\}$, for all i. Then $|\ell!|_2 = \ell - \sum_{i=0}^{s} \epsilon_i$.*

Proof. It is easy to see that

$$\begin{aligned}
|\ell!|_2 &= \left[\frac{\ell}{2}\right] + \left[\frac{\ell}{4}\right] + \left[\frac{\ell}{8}\right] + \cdots + 1 \\
&= \sum_{i=1}^{s} \epsilon_i 2^{i-1} + \sum_{i=2}^{s} \epsilon_i 2^{i-2} + \cdots + \sum_{i=s-1}^{s} \epsilon_i 2^{i-s+1} + \epsilon_s \\
&= \epsilon_1 + \epsilon_2 \sum_{i=0}^{1} 2^i + \epsilon_3 \sum_{i=0}^{2} 2^i + \cdots + \epsilon_s \sum_{i=0}^{s-1} 2^i \\
&= \epsilon_0(2^0 - 1) + \epsilon_1(2^1 - 1) + \epsilon_2(2^2 - 1) + \cdots + \epsilon_s(2^s - 1) \\
&= \ell - \sum_{i=0}^{s} \epsilon_i .
\end{aligned}$$

2.3. *Suppose $k = m2^{s+1} - 1$, with $s \geq 1$ and m odd. Then:*

(1) *If $1 \leq \ell < 2^s$, then $\binom{k+\ell}{2\ell} \equiv 0 \pmod 2$.*
(2) *If $1 \leq \ell < 2^s$, then $\binom{k+\ell}{2\ell+1} \equiv 0 \pmod 2$.*
(3) $\binom{k+2^s}{2^{s+1}} \equiv 1 \pmod 2$.
(4) $\binom{2k-2^s}{2^s} \equiv 0 \pmod 2$.
(5) $\binom{2k-2^s}{2^s-2} \equiv 1 \pmod 2$.
(6) $\binom{2k-2^s+1}{2^s-1} \equiv 1 \pmod 2$.

Proof. For (1) note that by comparing 2−parts of factors we have

$$\binom{k+\ell}{2\ell} \equiv \frac{\left\{\prod_{i=\ell-1}^{1}((k+1)+i)\right\} \cdot (k+1) \cdot \left\{\prod_{i=1}^{\ell}((k+1)-i)\right\}}{2^\ell \cdot \ell!} \pmod 2.$$

Since $k+1 = m2^{s+1}$ for $\ell \leq 2^s$, we get

$$\binom{k+\ell}{2\ell} \equiv \frac{(\ell-1)! \cdot 2^{s+1} \cdot \ell!}{2^\ell \cdot \ell!} \pmod 2$$

hence

$$\begin{aligned}
\left|\binom{k+\ell}{2\ell}\right|_2 &= \{|(\ell-1)!|_2 + s + 1 + |\ell!|_2\} - (\ell + |\ell!|_2) \\
&= |(\ell-1)!|_2 + s + 1 - \ell.
\end{aligned}$$

If $\ell < 2^s$, then $\ell - 1 < 2^s - 1$, so if we write $\ell - 1 = \sum_{i=0}^{s-1} \epsilon_i 2^i$, we see that $\sum_{i=0}^{s-1} \epsilon_i < s$. Thus, by 2.2, $|(\ell-1)!|_2 > \ell - 1 - s$, so $|\binom{k+\ell}{2\ell}|_2 > \ell - 1 - s + s + 1 - \ell = 0$. This shows (1). In (3), $\ell = 2^s$, so, by 2.2, $|(\ell-1)!|_2 = \ell - 1 - s$, thus $= |\binom{k+2^s}{2^{s+1}}|_2 = 0$.

For (2), note that

$$\binom{k+\ell}{2\ell+1} \equiv (k-\ell)\binom{k+\ell}{2\ell} \pmod 2.$$

Hence (2) following from (1).

We proceed with the proof of (4) and (5).

$$\begin{aligned}\binom{2k-2^s}{2^s} &\equiv \frac{\prod_{i=0}^{2^s-1}((m2^{s+2}-2^s-2)-i)}{2^s!} \equiv \frac{\prod_{i=0}^{2^s-1}(2^s+i+2)}{2^s!}\\ &\equiv \frac{2\cdot 2^s!}{2^s!} \equiv 0 \pmod 2,\end{aligned}$$

and as above,

$$\begin{aligned}\binom{2k-2^s}{2^s-2} &\equiv \frac{\prod_{i=0}^{2^s-3}((m2^{s+2}-2^s-2)-i)}{(2^s-2)!}\\ &\equiv \frac{\prod_{i=0}^{2^s-3}(2^s+i+2)}{(2^s-2)!} \equiv 1 \pmod 2.\end{aligned}$$

Finally, for (6), note that

$$\begin{aligned}\binom{2k-2^s+1}{2^s-1} &\equiv \frac{\prod_{i=0}^{2^s-2}((m2^{s+2}-2^s-1)-i)}{(2^s-1)!}\\ &\equiv \frac{\prod_{i=0}^{2^s-2}(2^s+i+1)}{(2^s-1)!} \equiv 1 \pmod 2.\end{aligned}$$

2.4. *Suppose $n = 2k$ and let $\tau : SL_n(\mathbb{F}) \to SL_n(\mathbb{F})$ be the automorphism defined in 1.4.4. Let $a_i, b_i \in SL_k(\mathbb{F})$ and suppose* $\mathrm{diag}\,(a_i, b_i^{-1}) \in \mathrm{Fix}\,(\tau)$, *$i = 1, 2$. Then for $\epsilon \in \{1, -1\}$, $F_{a_1^t a_2^\epsilon}[\lambda] = F_{b_1^t b_2^\epsilon}[\lambda]$.*

Proof. By 1.12.5, $b_i = J_k(a_i)^t J_k^t$. Hence, $b_1^t b_2 = J_k a_1 J_k^t J_k(a_2)^t J_k^t$. Recall now that $J_k^t = J_k^{-1}$. Hence $b_1^t b_2$ is conjugate to $a_1 a_2^t$, so $F_{a_1^t a_2} = F_{b_1^t b_2}$. Also $b_1^t b_2^{-1} = J_k a_1 J_k^t J_k(a_2^{-1})^t J_k^t$. Again we see that $b_1^t b_2^{-1}$ is conjugate to $a_1(a_2^{-1})^t$. Hence $F_{a_1^t a_2^{-1}} = F_{b_1^t b_2^{-1}}$.

2.5. *Let $m \geq 1$ and let $x = a_m$ or b_m. Then the characteristic polynomial of $x^t x^{-1}, x^{-1}x^t$, and $x(x^t)^{-1}$ is*

$$Q_m[\lambda] = \lambda^m - \lambda^{m-1} + \lambda^{m-2} - \cdots + (-1)^m.$$

Proof. First note that, by 1.11, $\operatorname{diag}(a_m, b_m^{-1}) \in \operatorname{Fix}(\tau)$, where $\tau : SL_{2m}(\mathbb{F}) \to SL_{2m}(\mathbb{F})$ is as defined in 1.4.4. Hence by 2.4,

(i) $$F_{a_m^t a_m^{-1}} = F_{b_m^t b_m^{-1}}.$$

Next, note that $x^t x^{-1}$ and $x^{-1}x^t$ are conjugate in $GL(m,\mathbb{F})$ and $x(x^t)^{-1}$, and $(x^t)^{-1}x$ are conjugate in $GL(m,\mathbb{F})$, so it suffices to show the lemma for $x^t x^{-1}$ and $x(x^t)^{-1}$. Now, by 2.7.1 (ahead), since $x(x^t)^{-1} = (x^t x^{-1})^{-1}$, $F_{x(x^t)^{-1}}[\lambda] = (-1)^m \lambda^m F_{x^t x^{-1}}[\lambda^{-1}]$, so if $F_{x^t x^{-1}}[\lambda] = Q_m[\lambda]$, then also $F_{x(x^t)^{-1}}[\lambda] = Q_m[\lambda]$. By (i), it remains to show that $Q_m[\lambda] = F_{a_m^t a_m^{-1}}[\lambda]$. Note now that,

$$a_m^t a_m^{-1} = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot & \bar{1} & 1 & \bar{1} & 1 \end{bmatrix}$$

and hence $F_{a_m^t a_m^{-1}}[\lambda] = Q_m[\lambda]$.

2.6. *Let $m \geq 1$. Then:*

(1) *For $x = a_m$ or b_m, $F_{x^t x}[\lambda] = F_{x x^t}[\lambda] = F_m[\lambda]$.*
(2) *For $m \geq 3$, $F_m = (\lambda - 2)F_{m-1} - F_{m-2}$, $F_m = (\lambda - 1)G_{m-1} - G_{m-2}$ and $G_m = (\lambda - 2)G_{m-1} - G_{m-2}$.*
(3) *$G_m[\lambda]$ is the characteristic polynomial of the $m \times m$ matrices*

$$y_m = \begin{bmatrix} 2 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & 2 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 2 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & 1 & 2 & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & 1 & 2 & 1 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 & 2 \end{bmatrix} \quad \textit{and}$$

$$z_m = \begin{bmatrix} 2 & \bar{1} & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \bar{1} & 2 & \bar{1} & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \bar{1} & 2 & \bar{1} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & \bar{1} & 2 & \bar{1} & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & \bar{1} & 2 & \bar{1} \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \bar{1} & 2 \end{bmatrix}.$$

(4) $F_m[\lambda] = \sum_{\ell=0}^{m}(-1)^{m+\ell}\binom{m+\ell}{2\ell}\lambda^\ell$.
(5) $G_m[\lambda] = \sum_{\ell=0}^{m}(-1)^{m+\ell}\binom{m+\ell+1}{2\ell+1}\lambda^\ell$.
(6) *Let* $\gamma \in \mathbb{F}$ *and suppose that for some* $\ell \geq 2$, $F_\ell[\gamma] = 0$. *Then* $G_{\ell-1}[\gamma] \neq 0$.

Proof. For (1), we already observed (using 1.11) that $\operatorname{diag}(a_m, b_m^{-1}) \in \operatorname{Fix}(\tau)$ and (1) follows from 2.4, and since, by definition, $F_m = F_{a_m^t a_m}$. Next, by definition $G_m = F_{y_m}$ (y_m as in (3)). Observe now that

$$a_m^t a_m = \begin{bmatrix} 2 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & 2 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 2 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & 1 & 2 & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & 1 & 2 & 1 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 & 1 \end{bmatrix}.$$

Now $F_m = \det(\lambda I_m - a_m^t a_m)$. Developing $\det(\lambda I_m - a_m^t a_m)$ using the first row, we easily get that for $m \geq 3$, $F_m = (\lambda - 2)F_{m-1} - F_{m-2}$. Developing $\det(\lambda I_m - a_m^t a_m)$ using the last row, we easily get $F_m = (\lambda-1)G_{m-1} - G_{m-2}$. Also developing $\det(\lambda I_m - y_m)$ using the first row gives $G_m = (\lambda-2)G_{m-1} - G_{m-2}$ and (2) is proved.

For (3), note that z_m is obtained from y_m by conjugating by $\operatorname{diag}(1, -1, 1, -1, \dots, (-1)^{m+1})$, so $F_{z_m}[\lambda] = F_{y_m}[\lambda] = G_m[\lambda]$.

To prove (4) and (5), note that $F_1 = \lambda - 1$, $F_2 = \lambda^2 - 3\lambda + 1$ and $G_1 = \lambda - 2$, $G_2 = \lambda^2 - 4\lambda + 3$. So (4) and (5) are the characteristic polynomials when $m = 1, 2$. Then, using (2), for $m \geq 3$, $\alpha(F_m, 0) = -2\alpha(F_{m-1}, 0) - \alpha(F_{m-2}, 0)$ and for $1 \leq \ell \leq m$, $\alpha(F_m, \ell) = \alpha(F_{m-1}, \ell - 1) - 2\alpha(F_{m-1}, \ell) - \alpha(F_{m-2}, \ell)$. The same equalities hold if we replace F by G. We must show that for $m \geq 3$.

(i) $$(-1)^m = -2(-1)^{m-1} - (-1)^{m-2}$$

(ii) $$(-1)^m\binom{m+1}{1} = -2(-1)^{m-1}\binom{m}{1} - (-1)^{m-2}\binom{m-1}{1}$$

(iii) $$\begin{aligned}(-1)^{m+\ell}\binom{m+\ell}{2\ell} &= (-1)^{m-1+\ell-1} \cdot \binom{m+\ell-2}{2\ell-2} \\ &\quad - 2(-1)^{m-1+\ell} \cdot \binom{m+\ell-1}{2\ell} \\ &\quad - (-1)^{m-2+\ell} \cdot \binom{m+\ell-2}{2\ell}\end{aligned}$$

(iv)
$$(-1)^{m+\ell}\cdot\binom{m+\ell+1}{2\ell+1} = (-1)^{m-1+\ell-1}\cdot\binom{m+\ell-1}{2\ell-1} - 2(-1)^{m-1+\ell}\cdot\binom{m+\ell}{2\ell+1} - (-1)^{m-2+\ell}\cdot\binom{m+\ell-1}{2\ell+1}.$$

For (i), note that $-2(-1)^{m-1} - (-1)^{m-2} = 2(-1)^m - (-1)^m$. For (ii), note that $-2(-1)^{m-1}\binom{m}{1} - (-1)^{m-2}\binom{m-1}{1} = 2(-1)^m m - (-1)^m(m-1) = (-1)^m(m+1)$.

For (iii) we have

$$\begin{aligned}
&(-1)^{m-1+\ell-1}\cdot\binom{m+\ell-2}{2\ell-2} - 2(-1)^{m-1+\ell}\cdot\binom{m+\ell-1}{2\ell} \\
&\quad - (-1)^{m-2+\ell}\cdot\binom{m+\ell-2}{2\ell} \\
&= (-1)^{m+\ell}\left\{\binom{m+\ell-2}{2\ell-2} + 2\binom{m+\ell-1}{2\ell} - \binom{m+\ell-2}{2\ell}\right\}.
\end{aligned}$$

Note now that

$$\begin{aligned}
&\binom{m+\ell-2}{2\ell-2} - \binom{m+\ell-2}{2\ell} \\
&= \binom{m+\ell-2}{2\ell-2} + \binom{m+\ell-2}{2\ell-1} - \binom{m+\ell-2}{2\ell-1} - \binom{m+\ell-2}{2\ell} \\
&= \binom{m+\ell-1}{2\ell-1} - \binom{m+\ell-1}{2\ell}.
\end{aligned}$$

Thus

$$\begin{aligned}
&(-1)^{m+\ell}\left\{\binom{m+\ell-2}{2\ell-2} + \binom{m+\ell-1}{2\ell} - \binom{m+\ell-2}{2\ell}\right\} \\
&= (-1)^{m+\ell}\left\{\binom{m+\ell-1}{2\ell-1} + \binom{m+\ell-1}{2\ell}\right\} \\
&= (-1)^{m+\ell}\cdot\binom{m+\ell}{2\ell}
\end{aligned}$$

and (iii) is proved.

For (iv) we have

$$\begin{aligned}&(-1)^{m-1+\ell-1}\cdot\binom{m+\ell-1}{2\ell-1}-2(-1)^{m-1+\ell}\cdot\binom{m+\ell}{2\ell+1}\\&\quad-(-1)^{m-2+\ell}\cdot\binom{m+\ell-1}{2\ell+1}\\&=(-1)^{m+\ell}\left\{\binom{m+\ell-1}{2\ell-1}+2\binom{m+\ell}{2\ell+1}-\binom{m+\ell-1}{2\ell+1}\right\}\end{aligned}$$

and as in the previous paragraph of the proof we get (iv). This shows (4) and (5).

Suppose that $F_\ell[\gamma]=0=G_{\ell-1}[\gamma]$, for some $\ell\geq 2$, then, by (2), also $G_{\ell-2}[\gamma]=0$. Then, using (2), we see that $G_m[\gamma]=0$, for all $1\leq m\leq\ell$. In particular, $G_1[\gamma]=0=G_2[\gamma]$, so $\gamma=2$ and $0=2^2-4\cdot 2+3=-1$, a contradiction.

2.7. *Let $h,g\in SL_n(\mathbb{F})$ and let $Q[\lambda]=F_g$. Then:*

(1) $\overline{Q}=(-1)^n\lambda^nQ[\lambda^{-1}]$. *In particular, for all* $0\leq\ell\leq n$, $\alpha(\overline{Q},\ell)=(-1)^n\alpha(Q,n-\ell)$.
(2) $F_{hg}[\lambda]=F_{gh}[\lambda]=\det(\lambda h^{-1}-g)$.
(3) *Suppose $\ell,m\geq 1$ are integers and $\epsilon\in\{1,-1\}$. Suppose $h^{-1}=\operatorname{diag}(I_{\ell-1},s^{-1},I_{m-1})$, where s is a $(2+\epsilon)\times(2+\epsilon)$ matrix. Then $F_{hg}=\det(r+(\lambda I-g))$, where $r=\operatorname{diag}(0_{\ell-1},\lambda s^{-1}-\lambda I_{2+\epsilon},0_{m-1})$.*

Proof. Set $I=I_n$. Then $F_{g^{-1}}=\det(\lambda I-g^{-1})=\det\{-\lambda I(\lambda^{-1}I-g)g^{-1}\}=(-\lambda)^n\det(\lambda^{-1}I-g)=(-1)^n\lambda^nQ[\lambda^{-1}]$.

For (2), we have $\det(\lambda I-gh)=\det\{(\lambda h^{-1}-g)h\}=\det(\lambda h^{-1}-g)$. Finally, for (3), $\det(\lambda h^{-1}-g)=\det(\lambda h^{-1}-\lambda I+\lambda I-g)=\det(r+\lambda I-g)$, because $r=\lambda h^{-1}-\lambda I$.

2.8. *Let $\ell,m\geq 1$ be two integers such that $\ell+m=2k$. Let $\mathfrak{A}\in M_\ell(\mathfrak{R})$ and $\mathfrak{B}\in M_m(\mathfrak{R})$. If $\epsilon=0$, let $g=\operatorname{diag}(\mathfrak{A},\mathfrak{B})$, while if $\epsilon=1$, let $g=\operatorname{diag}(\mathfrak{A},\mu,\mathfrak{B})$, with $0\neq\mu\in\mathfrak{R}$. Let f be the following $(2+\epsilon)\times(2+\epsilon)$ matrix over $\mathfrak{R}$*

$$f=\begin{bmatrix}\alpha&\beta\\\gamma&\delta\end{bmatrix}\qquad\textit{when }\ \epsilon=0,$$

$$f=\begin{bmatrix}\alpha_{11}&\alpha_{12}&\alpha_{13}\\\alpha_{21}&\alpha_{22}&\alpha_{23}\\\alpha_{31}&\alpha_{32}&\alpha_{33}\end{bmatrix}\qquad\textit{when }\ \epsilon=1.$$

Let $r=\operatorname{diag}(0_{\ell-1},f,0_{m-1})$. Then:

(1) *If* $\epsilon = 0$, *then*

$$\begin{aligned}\det(r+g) &= \det(\mathfrak{A})\det(\mathfrak{B}) + \delta\det(\mathfrak{A})\det(M_{1,1}(\mathfrak{B}))\\ &\quad + \alpha\det(M_{\ell,\ell}(\mathfrak{A}))\det(\mathfrak{B})\\ &\quad + \det\begin{bmatrix}\alpha & \beta\\ \gamma & \delta\end{bmatrix}\det(M_{\ell,\ell}(\mathfrak{A}))\det(M_{1,1}(\mathfrak{B})).\end{aligned}$$

(2) *If* $\epsilon = 1$, *then*

$$\begin{aligned}\det(r+g) &= (\alpha_{22}+\mu)\det(\mathfrak{A})\det(\mathfrak{B})\\ &\quad + \det\begin{bmatrix}\alpha_{22}+\mu & \alpha_{23}\\ \alpha_{32} & \alpha_{33}\end{bmatrix}\det(\mathfrak{A})\det(M_{1,1}(\mathfrak{B}))\\ &\quad + \det\begin{bmatrix}\alpha_{11} & \alpha_{12}\\ \alpha_{21} & \alpha_{22}+\mu\end{bmatrix}\det(M_{\ell,\ell}(\mathfrak{A}))\det(\mathfrak{B})\\ &\quad + \det\begin{bmatrix}\alpha_{11} & \alpha_{12} & \alpha_{13}\\ \alpha_{21} & \alpha_{22}+\mu & \alpha_{23}\\ \alpha_{31} & \alpha_{32} & \alpha_{33}\end{bmatrix}\det(M_{\ell,\ell}(\mathfrak{A}))\det(M_{1,1}(\mathfrak{B})).\end{aligned}$$

Proof. (1) is proved by expanding $\det(r+g)$ along row $\ell+1$. For (2), expanding $\det(r+g)$ along the $(\ell+1)$-row, we get

$$\text{(i)}\quad \det(r+g) = -\alpha_{21}\det(r_1+g_1) + (\alpha_{22}+\mu)\det(r_2+g_2) - \alpha_{23}\det(r_3+g_3)$$

where $r_1 = \operatorname{diag}\left(0_{\ell-1}, \begin{bmatrix}\alpha_{12} & \alpha_{13}\\ \alpha_{32} & \alpha_{33}\end{bmatrix}, 0_{m-1}\right)$, $g_1 = \operatorname{diag}(\mathfrak{A}_1, \mathfrak{B})$, and $\mathfrak{A}_1$ is obtained from $\mathfrak{A}$ by replacing the last column by a column of zeros. $r_2 = \operatorname{diag}\left(0_{\ell-1}, \begin{bmatrix}\alpha_{11} & \alpha_{13}\\ \alpha_{31} & \alpha_{33}\end{bmatrix}, 0_{m-1}\right)$, and $g_2 = \operatorname{diag}(\mathfrak{A}, \mathfrak{B})$. $r_3 = \operatorname{diag}\left(0_{\ell-1}, \begin{bmatrix}\alpha_{11} & \alpha_{12}\\ \alpha_{31} & \alpha_{32}\end{bmatrix}, 0_{m-1}\right)$, $g_3 = \operatorname{diag}(\mathfrak{A}, \mathfrak{B}_1)$, and $\mathfrak{B}_1$ is obtained from $\mathfrak{B}$ by replacing the first column by a column of zeros. Notice now that $\det(\mathfrak{A}_1) = 0 = \det(\mathfrak{B}_1)$ and $\det(M_{\ell,\ell}(\mathfrak{A}_1)) = \det(M_{\ell,\ell}(\mathfrak{A}))$, while $\det(M_{1,1}(\mathfrak{B}_1)) = \det(M_{1,1}(\mathfrak{B}))$. Now, by (1), we get

$$\text{(ii)}\quad \begin{aligned}\det(r_1+g_1) &= \alpha_{12}\det(M_{\ell,\ell}(\mathfrak{A}))\det(\mathfrak{B})\\ &\quad + \det\begin{bmatrix}\alpha_{12} & \alpha_{13}\\ \alpha_{32} & \alpha_{33}\end{bmatrix}\det(M_{\ell,\ell}(\mathfrak{A}))\det(M_{1,1}(\mathfrak{B})).\end{aligned}$$

(iii) $$\det(r_2+g_2) = \det(\mathfrak{A})\det(\mathfrak{B}) + \alpha_{33}\det(\mathfrak{A})\det(M_{1,1}(\mathfrak{B})) + \alpha_{11}\det(M_{\ell,\ell}(\mathfrak{A}))\det(\mathfrak{B}) + \det\begin{bmatrix}\alpha_{11} & \alpha_{13}\\ \alpha_{31} & \alpha_{33}\end{bmatrix}\det(M_{\ell,\ell}(\mathfrak{A}))\det(M_{1,1}(\mathfrak{B})).$$

(iv) $$\det(r_3+g_3) = \alpha_{32}\det(\mathfrak{A})\det(M_{1,1}(\mathfrak{B})) + \det\begin{bmatrix}\alpha_{11} & \alpha_{12}\\ \alpha_{31} & \alpha_{32}\end{bmatrix}\det(M_{\ell,\ell}(\mathfrak{A}))\det(M_{1,1}(\mathfrak{B})).$$

Note now that (2) follows from (i)-(iv).

2.9. *Let $\ell, m \geq 1$ be two integers such that $\ell + m = 2k$. Let $A \in M_\ell(\mathbb{F})$ and $B \in M_m(\mathbb{F})$. Let $g = \operatorname{diag}(A, B)$. Let $s \in GL_2(\mathbb{F})$ such that $s^{-1} = \begin{bmatrix}\beta_{11} & \beta_{12}\\ \beta_{21} & \beta_{22}\end{bmatrix}$. Let $h = \operatorname{diag}(I_{\ell-1}, s, I_{m-1})$. Then*

$$F_{hg} = F_AF_B + (\beta_{22}-1)\lambda F_AF_{M_{1,1}(B)} + (\beta_{11}-1)\lambda F_{M_{\ell,\ell}(A)}F_B + \det\begin{bmatrix}(\beta_{11}-1)\lambda & \beta_{12}\lambda\\ \beta_{21}\lambda & (\beta_{22}-1)\lambda\end{bmatrix} F_{M_{\ell,\ell}(A)}F_{M_{1,1}(A)}.$$

Proof. First we mention, that, by definition, if R is a 1×1 matrix over $\mathbb{F}$, we always take $F_{M_{1,1}(R)} = 1$. Next note that $h^{-1} = \operatorname{diag}(I_{\ell-1}, s^{-1}, I_{m-1})$. By 2.7.3, $F_{gh} = \det(r + (\lambda I_n - g))$, where $r = \operatorname{diag}(0_{\ell-1}, \lambda s^{-1} - \lambda I_2, 0_{m-1})$. Note now that

$$\lambda s^{-1} - \lambda I_2 = \begin{bmatrix}(\beta_{11}-1)\lambda & \beta_{12}\lambda\\ \beta_{21}\lambda & (\beta_{22}-1)\lambda\end{bmatrix}$$

also,

$$\lambda I_n - g = \operatorname{diag}(\lambda I_\ell - A, \lambda I_m - B).$$

So if we set $\mathfrak{A} = \lambda I_\ell - A$ and $\mathfrak{B} = \lambda I_m - B$, then by 2.8.1,

$$\begin{aligned}&\det(r + (\lambda I - g))\\ &= \det(\mathfrak{A})\det(\mathfrak{B}) + (\beta_{22}-1)\lambda\det(\mathfrak{A})\det(M_{1,1}(\mathfrak{B}))\\ &\quad + (\beta_{11}-1)\lambda\det(M_{\ell,\ell}(\mathfrak{A}))\det(\mathfrak{B})\\ &\quad + \det\begin{bmatrix}(\beta_{11}-1)\lambda & \beta_{12}\lambda\\ \beta_{21}\lambda & (\beta_{22}-1)\lambda\end{bmatrix}\det(M_{\ell,\ell}(\mathfrak{A}))\det(M_{1,1}(\mathfrak{B})).\end{aligned}$$

The lemma follows.

2.10. *Let $g = \operatorname{diag}(A, 1, B)$, with $A, B \in M_k(\mathbb{F})$. Let $s \in SL_3(\mathbb{F})$ such that*

$$s^{-1} = \begin{bmatrix}\beta_{11} & \beta_{12} & \beta_{13}\\ \beta_{21} & \beta_{22} & \beta_{23}\\ \beta_{31} & \beta_{32} & \beta_{33}\end{bmatrix}.$$

Let $h = \operatorname{diag}(I_{k-1}, s, I_{k-1})$. *Then* $\alpha(F_{hg}, 1) = \alpha(R[\lambda], 1)$, *where*

$$R[\lambda] = (\beta_{22}\lambda - 1)F_A F_B - (\beta_{33} - 1)\lambda F_A F_{M_{1,1}(B)} - (\beta_{11} - 1)\lambda F_{M_{k,k}(A)} F_B.$$

Proof. We use 2.8.2, with $\ell = m = k$. First note that $h^{-1} = \operatorname{diag}(I_{k-1}, s^{-1}, I_{k-1})$. By 2.7.3, $F_{gh} = \det(r + (\lambda I - g))$, where

$$r = \operatorname{diag}(0_{k-1}, \lambda s^{-1} - \lambda I_3, 0_{k-1}).$$

Note now that

$$\lambda s^{-1} - \lambda I_3 = \begin{bmatrix} (\beta_{11} - 1)\lambda & \beta_{12}\lambda & \beta_{13}\lambda \\ \beta_{21}\lambda & (\beta_{22} - 1)\lambda & \beta_{23}\lambda \\ \beta_{31}\lambda & \beta_{32}\lambda & (\beta_{33} - 1)\lambda \end{bmatrix}$$

also, if we set $I = I_n$, then

$$\lambda I - g = \operatorname{diag}(\lambda I_k - A, \lambda - 1, \lambda I_k - B).$$

We use 2.8.2 with $\mathfrak{A} = \lambda I_k - A$, $\mathfrak{B} = \lambda I_k - B$ and $\mu = \lambda - 1$. The α_{ij} are given by the matrix $\lambda s^{-1} - \lambda I_3$ above. By 2.8.2

$$\begin{aligned} &\det(r + (\lambda I - g)) \\ &= (\beta_{22}\lambda - 1)\det(\mathfrak{A})\det(\mathfrak{B}) \\ &\quad + \det\begin{bmatrix} \beta_{22}\lambda - 1 & \beta_{23}\lambda \\ \beta_{32}\lambda & (\beta_{33} - 1)\lambda \end{bmatrix} \det(\mathfrak{A})\det(M_{1,1}(\mathfrak{B})) \\ &\quad + \det\begin{bmatrix} (\beta_{11} - 1)\lambda & \beta_{12}\lambda \\ \beta_{21}\lambda & \beta_{22}\lambda - 1 \end{bmatrix} \det(M_{k,k}(\mathfrak{A}))\det(\mathfrak{B}) \\ &\quad + \det\begin{bmatrix} (\beta_{11} - 1)\lambda & \beta_{12}\lambda & \beta_{13}\lambda \\ \beta_{21}\lambda & \beta_{22}\lambda - 1 & \beta_{23}\lambda \\ \beta_{31}\lambda & \beta_{32}\lambda & (\beta_{33} - 1)\lambda \end{bmatrix} \det(M_{k,k}(\mathfrak{A}))\det(M_{1,1}(\mathfrak{B})) \end{aligned}$$

so we see that the only expressions in $\det(r + (\lambda I - g))$ which contribute to the coefficient of λ in $\det(r + (\lambda I - g))$ are

$$\begin{aligned} (\beta_{22}\lambda - 1)\det(\mathfrak{A})\det(\mathfrak{B}) - (\beta_{33} - 1)\lambda\det(\mathfrak{A})\det(M_{1,1}(\mathfrak{B})) \\ - (\beta_{11} - 1)\lambda\det(M_{k,k}(\mathfrak{A}))\det(\mathfrak{B}) \end{aligned}$$

because the other expressions are in $\lambda^2\mathbb{F}[\lambda]$. This shows the lemma.

2.11. *Let* $m \geq 2$ *be an integer and let* $c, d \in SL_m(\mathbb{F})$ *be two unipotent elements such that* c *is lower triangular and* d *is upper triangular. Let* $x \in SL_m(\mathbb{F})$. *Then:*

(1) $M_{\ell,\ell}(dx) = M_{\ell,\ell}(d)M_{\ell,\ell}(x)$, *for* $\ell \in \{1, (1,2)\}$.
(2) $M_{\ell,\ell}(xc) = M_{\ell,\ell}(x)M_{\ell,\ell}(c)$, *for* $\ell \in \{1, (1,2)\}$.
(3) $M_{m,m}(cx) = M_{m,m}(c)M_{m,m}(x)$ *and* $M_{m,m}(xd) = M_{m,m}(x)M_{m,m}(d)$.
(4) $M_{\ell,\ell}(y^{-1}) = \{M_{\ell,\ell}(y)\}^{-1}$, *for* $y \in \{c, d\}$ *and* $\ell \in \{1, m, (1,2)\}$.

Proof. (1), (2) and (3) are obvious and (4) follows from them.

2.12. *Let $m \geq 3$, $\beta_1, \beta_2, \dots, \beta_m$, $\gamma_1, \gamma_2, \dots, \gamma_m \in \mathbb{F}^*$. For $1 \leq i \leq 3$, let*

$$B_i := b_{m+2-i}(\beta_i, \dots, \beta_m) \quad \text{and} \quad C_i := b_{m+2-i}(\gamma_i, \dots, \gamma_m).$$

Then:

(1) $F_{C_1^t B_1} = (\lambda - 1) F_{C_2^t B_2} - \beta_1 \gamma_1 \lambda F_{M_{1,1}(B_2 C_2^t)}$.
(2) $F_{(C_1^t B_1)^{-1}} = \{(1 + \beta_1 \gamma_1)\lambda - 1\} F_{(C_2^t B_2)^{-1}} - \beta_1 \gamma_1 \lambda^2 F_{(C_3^t B_3)^{-1}}$.
(3) $F_{C_1^t B_1^{-1}} = (\lambda - 1) F_{C_2^t B_2^{-1}} + \beta_1 \gamma_1 \lambda F_{C_3^t B_3^{-1}}$.
(4) *If* $B_2 = C_2 = b_m$, $F_{C_1^t B_1} = (\lambda - 1) F_m - \beta_1 \gamma_1 \lambda G_{m-1}$.
(5) *If* $B_2 = C_2 = b_m$, *then* $F_{C_1^t B_1^{-1}} = (\lambda - 1) Q_m + \beta_1 \gamma_1 \lambda Q_{m-1}$.

Proof. First note that (4) and (5) follow from (1) and (3) respectively, since, if $B_2 = C_2 = b_m$, then, by 2.6, $F_{C_2^t B_2} = F_m$ and, by 2.5, $F_{C_2^t B_2^{-1}} = Q_m$, $F_{C_3^t B_3^{-1}} = Q_{m-1}$ and we leave it for the reader to verify that $F_{M_{1,1}(B_2 C_2^t)} = G_{m-1}$.

To prove (1), (2) and (3), let $u = u_1^{m+1}(-\beta_1)$ and $w = u_1^{m+1}(-\gamma_1)$. Note first that $B_1 = u \operatorname{diag}(1, B_2)$ and $C_1 = w \operatorname{diag}(1, C_2)$. Hence

$$\text{(i)} \qquad C_1^t B_1 = \operatorname{diag}(1, C_2^t) w^t u \operatorname{diag}(1, B_2)$$

$$\text{(ii)} \qquad (C_1^t B_1)^{-1} = \operatorname{diag}(1, B_2^{-1}) u^{-1} (w^t)^{-1} \operatorname{diag}(1, (C_2^t)^{-1})$$

$$\text{(iii)} \qquad C_1^t B_1^{-1} = \operatorname{diag}(1, C_2^t) \operatorname{diag}(1, B_2^{-1}) w^t u^{-1}$$

where (iii) follows from the fact that $\operatorname{diag}(1, B_2^{-1})$ and w^t commute.

For (1), (2) and (3), given $S \in \{C_1^t B_1, C_1^t B_1^{-1}, (C_1^t B_1)^{-1}\}$, we find $g, h \in SL_{m+1}(\mathbb{F})$ and $B \in SL_m(\mathbb{F})$ (g, h and B depend on S) such that S is conjugate to hg, with $g = \operatorname{diag}(1, B)$ and $h^{-1} = \operatorname{diag}(s, I_{m-1})$. Then we use 2.9 (with $\ell = 1$ and $m = m$) to compute F_{hg}. Note that by 2.9 if $A \in M_1(\mathbb{F})$, $B \in M_m(\mathbb{F})$, then for $g = \operatorname{diag}(A, B)$ and $h^{-1} = \operatorname{diag}\left(\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}, I_{m-1}\right)$,

$$\text{(iv)} \qquad F_{hg} = F_A F_B + (\delta - 1)\lambda F_A F_{M_{1,1}(B)} + (\alpha - 1)\lambda F_B + \det \begin{bmatrix} (\alpha - 1)\lambda & \beta\gamma \\ \gamma\lambda & (\delta - 1)\lambda \end{bmatrix} F_{M_{1,1}(B)}.$$

In all cases we take $A = 1$.

(v) In (1), take $B = B_2 C_2^t$; in (2) take $B = (B_2 C_2^t)^{-1}$; in (3) take $B = C_2^t B_2^{-1}$.

Also

(vi) in (1), take $h^{-1} = (w^t u)^{-1} = \operatorname{diag}\left(\begin{bmatrix} 1 & \gamma_1 \\ \beta_1 & \beta_1\gamma_1 + 1 \end{bmatrix}, I_{m-1}\right)$;

in (2) take $h^{-1} = w^t u = \operatorname{diag}\left(\begin{bmatrix} 1 + \beta_1\gamma_1 & -\gamma_1 \\ -\beta_1 & 1 \end{bmatrix}, I_{m-1}\right)$;

in (3) take $h^{-1} = (w^t u^{-1})^{-1} = \operatorname{diag}\left(\begin{bmatrix} 1 & \gamma_1 \\ -\beta_1 & -\beta_1\gamma_1 + 1 \end{bmatrix}, I_{m-1}\right)$.

We now use (iv), (v) and (vi) to prove (1) (2) and (3).

In (1), taking $B = B_2C_2^t$, we get

$$\begin{aligned} F_{C_1^tB_1} &= (\lambda-1)F_B + \beta_1\gamma_1\lambda(\lambda-1)F_{M_{1,1}(B)} \\ &\quad + \det\begin{bmatrix} 0 & \gamma_1\lambda \\ \beta_1\lambda & \beta_1\gamma_1\lambda \end{bmatrix} F_{M_{1,1}(B)} \\ &= (\lambda-1)F_{B_2C_2^t} - \beta_1\gamma_1\lambda F_{M_{1,1}(B_2C_2^t)} \end{aligned}$$

also, in (3), taking $B = C_2^tB_2^{-1}$, we get

$$\begin{aligned} F_{C_1^tB_1^{-1}} &= (\lambda-1)F_B - \beta_1\gamma_1\lambda(\lambda-1)F_{M_{1,1}(B)} \\ &\quad + \det\begin{bmatrix} 0 & \gamma_1\lambda \\ -\beta_1\lambda & -\beta_1\gamma_1\lambda \end{bmatrix} F_{M_{1,1}(B)} \\ &= (\lambda-1)F_{C_2^tB_2^{-1}} + \beta_1\gamma_1\lambda F_{M_{1,1}(C_2^tB_2^{-1})}. \end{aligned}$$

Since $M_{1,1}(C_2^tB_2^{-1}) = C_3^tB_3^{-1}$, we get (3). Finally in (2), taking $B = (B_2C_2^t)^{-1}$, we get

$$\begin{aligned} F_{(C_1^tB_1)^{-1}} &= (\lambda-1)F_B + \beta_1\gamma_1\lambda F_B \\ &\quad + \det\begin{bmatrix} \beta_1\gamma_1\lambda & -\gamma_1\lambda \\ -\beta_1\lambda & 0 \end{bmatrix} F_{M_{1,1}(B)} \\ &= \{\lambda - 1 + \beta_1\gamma_1\lambda\}F_{(B_2C_2^t)^{-1}} - \beta_1\gamma_1\lambda^2 F_{M_{1,1}((B_2C_2^t)^{-1})}. \end{aligned}$$

Note however that $F_{(B_2C_2^t)^{-1}} = F_{(C_2^tB_2)^{-1}}$ and that, by 2.11.1, $M_{1,1}\{(B_2C_2^t)^{-1}\} = (B_3C_3^t)^{-1}$ and again $F_{(B_3C_3^t)^{-1}} = F_{(C_3^tB_3)^{-1}}$.

2.13. *Suppose $n=2k$. Let $\alpha \in \mathbb{F}^*$ and set $u = u_k^n(\alpha)$. Let $X = \operatorname{diag}(a_k, b_k^{-1})u$ and let H_n be the characteristic polynomial of X^tX. Then:*

(1) $$H_n = \overline{F}_k(F_k + \alpha^2\lambda G_{k-1}) - \alpha^2\lambda^2 G_{k-1}\overline{F}_{k-1}.$$

(2) $$\alpha(H_n, 1) = -\binom{k+1}{2} - (\alpha^2+2)k + 1.$$

Suppose $\alpha = 1$. Then:

(3) *If* $\operatorname{char}(\mathbb{F}) = 3$ *and* $k \equiv 0$ *or* $2 \pmod 3$*, then* $\alpha(H_n, 1) \neq 0$.
(4) *If* $\operatorname{char}(\mathbb{F}) = 2$ *and* $k \equiv 0$ *or* $1 \pmod 4$*, then* $\alpha(H_n, 1) \neq 0$.
(5) *If* $\operatorname{char}(F) = 2$ *and* $k \equiv -2$ *or* $3 \pmod 8$*, then* $\alpha(H_n, 2) \neq 0$.
(6) *If* $\operatorname{char}(F) = 2$ *and* $k \equiv 2 \pmod 8$*, then either* $\alpha(H_n, 4) \neq 0$ *or* $\alpha(H_n, 7) \neq 0$.
(7) *If* $\operatorname{char}(F) = 2$ *and* $k \equiv -1 \pmod 8$*, then* $\alpha(H_n, 2^s) = 1$*, where s is defined by $k = m2^{s+1} - 1$, with m odd.*

Proof. For (1), we'll use 2.9. But first we observe that

(i) $$X^tX = u^t\operatorname{diag}(a_k^ta_k, (b_k^t)^{-1}b_k^{-1})u.$$

Further, by definition and by 2.6.1,

(ii) $$F_{a_k^t a_k} = F_k \qquad F_{(b_k^t)^{-1}b_k^{-1}} = \overline{F}_k.$$

Also, by 2.11.1 and 2.11.4,

(iii) $$M_{1,1}((b_k^t)^{-1}b_k^{-1}) = (b_{k-1}^t)^{-1}b_{k-1}^{-1} \quad \text{so} \quad F_{M_{1,1}((b_k^t)^{-1}b_k^{-1})} = \overline{F}_{k-1}.$$

Finally observe that by definition and by the shape of $a_k^t a_k$

(iv) $$F_{M_{k,k}(a_k^t a_k)} = G_{k-1}.$$

Set $h = uu^t$. Of course $h = \operatorname{diag}(I_{k-1}, s, I_{k-1})$, with $s^{-1} = \begin{bmatrix} \alpha^2+1 & \bar{\alpha} \\ \bar{\alpha} & 1 \end{bmatrix}$. Note that, by (i), H_n is the characteristic polynomial of hg, with $g = \operatorname{diag}(A, B)$, $A = a_k^t a_k$ and $B = (b_k^t)^{-1}b_k^{-1}$. Thus by 2.9

$$\begin{aligned} H_n = F_{hg} &= F_A F_B + \alpha^2\lambda F_{M_{k,k}(A)} F_B \\ &\qquad + \det \begin{bmatrix} \alpha^2\lambda & \bar{\alpha}\lambda \\ \bar{\alpha}\lambda & 0 \end{bmatrix} F_{M_{k,k}(A)} F_{M_{1,1}(B)} \\ &= F_A F_B + \alpha^2\lambda F_{M_{k,k}(A)} F_B - \alpha^2\lambda^2 F_{M_{k,k}(A)} F_{M_{1,1}(B)}. \end{aligned}$$

Using (ii), (iii) and (iv) we see that (1) holds. Next, using 2.6 and 2.7,

$$\alpha(H_n, 1) = \alpha(\overline{F}_k, 0)\{\alpha(F_k, 1) + \alpha^2\alpha(G_{k-1}, 0)\} + \alpha(F_k, 0)\alpha(\overline{F}_k, 1)$$

$$\alpha(\overline{F}_k, 0) = (-1)^k = \alpha(F_k, 0), \quad \alpha(G_{k-1}, 0) = (-1)^{k-1}\binom{k}{1}$$

$$\alpha(F_k, 1) = (-1)^{k+1}\binom{k+1}{2}, \quad \alpha(\overline{F}_k, 1) = (-1)^k(1-2k).$$

Thus

$$\begin{aligned} \alpha(H_n, 1) &= (-1)^k \left\{ (-1)^{k+1}\binom{k+1}{2} + \alpha^2(-1)^{k-1}\binom{k}{1} \right\} \\ &\qquad + (-1)^k(-1)^k(1-2k) \\ &= -\binom{k+1}{2} - \binom{k}{1}\alpha^2 - 2k + 1 \\ &= -\binom{k+1}{2} - (\alpha^2+2)k + 1. \end{aligned}$$

This shows (2). For the remainder of the proof we assume that $\alpha = 1$. Suppose first that $\operatorname{char}(\mathbb{F}) = 3$. By (2), $\alpha(H_n, 1) = -\binom{k+1}{2} + 1$. Thus if $k \equiv 0$ or $2 \pmod 3$, $\alpha(H_n, 1) \neq 0$ and (3) is proved.

So suppose that $\operatorname{char}(\mathbb{F}) = 2$. By (2), $\alpha(H_n, 1) = \binom{k+1}{2} + k + 1$. Hence if $k \equiv 0$ or $1 \pmod 4$, $\alpha(H_n, 1) = 1$ and (4) is proved. Recall from 2.6 and 2.7

that

(*)

$$\begin{aligned} F_k[\lambda] &= 1 + \binom{k+1}{2}\lambda + \binom{k+2}{4}\lambda^2 + \binom{k+3}{6}\lambda^3 + \binom{k+4}{8}\lambda^4 + \cdots \\ \overline{F}_k[\lambda] &= 1 + \binom{2k-1}{1}\lambda + \binom{2k-2}{2}\lambda^2 + \binom{2k-3}{3}\lambda^3 + \binom{2k-4}{4}\lambda^4 + \cdots \\ \overline{F}_{k-1}[\lambda] &= 1 + \binom{2k-3}{1}\lambda + \binom{2k-4}{2}\lambda^2 + \binom{2k-5}{3}\lambda^3 + \binom{2k-6}{4}\lambda^4 + \cdots \\ G_{k-1} &= k + \binom{k+1}{3}\lambda + \binom{k+2}{5}\lambda^2 + \binom{k+3}{7}\lambda^3 + \binom{k+4}{9}\lambda^4 + \cdots . \end{aligned}$$

Suppose first that $k \equiv -2 \pmod 8$. Using (*), note that $\overline{F}_k \equiv 1 + \lambda + \lambda^2 \pmod{(\lambda^3)}$, $F_k \equiv 1 + \lambda \pmod{(\lambda^3)}$ and $G_{k-1} \equiv \lambda \pmod{(\lambda^2)}$. Hence modulo the ideal (λ^3), $\overline{F}_k(F_k + \lambda G_{k-1}) - \lambda^2 G_{k-1}\overline{F}_{k-1} \equiv (1+\lambda+\lambda^2)(1+\lambda+\lambda^2) \equiv 1+\lambda^2$. Thus $\alpha(H_n, 2) \neq 0$.

Suppose $k \equiv 3 \pmod 8$. Then by (*), $\overline{F}_k \equiv 1+\lambda \pmod{(\lambda^3)}$, $F_k \equiv 1+\lambda^2 \pmod{(\lambda^3)}$, $G_{k-1} \equiv 1 \pmod{(\lambda^2)}$ and $\overline{F}_{k-1} \equiv 1 \pmod{(\lambda)}$. Hence, modulo the ideal (λ^3), $\overline{F}_k(F_k + \lambda G_{k-1}) - \lambda^2 G_{k-1}\overline{F}_{k-1} \equiv (1+\lambda)(1+\lambda^2+\lambda) + \lambda^2 \equiv 1+\lambda^2$. This completes the proof of (5).

Suppose $k = 8m+2$. Note that $\binom{k+1}{2} \equiv 1 \pmod 2$, $\binom{k+2}{4} \equiv \frac{4\cdot 2}{4\cdot 2} \equiv 1 \pmod 2$, $\binom{k+3}{6} \equiv \frac{4\cdot 2\cdot(k-2)}{2\cdot 4\cdot 2} \equiv 0 \pmod 2$, $\binom{k+4}{8} \equiv \frac{2\cdot 4\cdot 2\cdot(k-2)}{8\cdot 2\cdot 4\cdot 2} \equiv m \pmod 2$, $\binom{k+5}{10} \equiv \frac{2\cdot 4\cdot 2\cdot(k-2)\cdot 2}{2\cdot 8\cdot 2\cdot 4\cdot 2} \equiv m \pmod 2$. $\binom{k+6}{12} \equiv \frac{(k+6)2\cdot 4\cdot 2\cdot(k-2)\cdot 2}{4\cdot 2\cdot 8\cdot 2\cdot 4\cdot 2} \equiv \frac{(k+6)\cdot(k-2)}{4\cdot 8} \equiv 0 \pmod 2$, and similarly, $\binom{k+7}{14} \equiv 0 \pmod 2$. Hence, by (*),

$$F_k \equiv 1 + \lambda + \lambda^2 + m\lambda^4 + m\lambda^5 \pmod{(\lambda^8)}.$$

Next, $\binom{2k-1}{1} \equiv 1 \pmod 2$, $\binom{2k-2}{2} \equiv 1 \pmod 2$, $\binom{2k-3}{3} \equiv 0 \pmod 2$, $\binom{2k-4}{4} \equiv 0 \pmod 2$ and $\binom{2k-5}{5} \equiv \frac{2\cdot 4}{4\cdot 2} \equiv 1 \pmod 2$, $\binom{2k-6}{6} \equiv \frac{2\cdot 4\cdot 2}{2\cdot 4\cdot 2} \equiv 1 \pmod 2$, $\binom{2k-7}{7} \equiv 0 \pmod 2$. Hence, by (*),

$$\overline{F}_k = 1 + \lambda + \lambda^2 + \lambda^5 + \lambda^6 \pmod{(\lambda^8)}.$$

Next, $\binom{2k-3}{1} \equiv 1 \pmod 2$, $\binom{2k-4}{2} \equiv 0 \pmod 2$, $\binom{2k-5}{3} \equiv 1 \pmod 2$, $\binom{2k-6}{4} \equiv 1 \pmod 2$, $\binom{2k-7}{5} \equiv 1 \pmod 2$. Hence, by (*),

$$\overline{F}_{k-1} = 1 + \lambda + \lambda^3 + \lambda^4 + \lambda^5 \pmod{(\lambda^6)}.$$

Finally, $\binom{k}{1} \equiv 0 \pmod 0 \pmod 2$, $\binom{k+1}{3} \equiv 1 \pmod 2$, $\binom{k+2}{5} \equiv \frac{4\cdot 2\cdot(k-2)}{2\cdot 4\cdot 2} \equiv 0 \pmod 2$, $\binom{k+3}{7} \equiv \frac{4\cdot 2\cdot(k-2)}{2\cdot 4\cdot 2} \equiv 0 \pmod 2$, $\binom{k+4}{9} \equiv \frac{2\cdot 4\cdot 2\cdot(k-2)\cdot 2}{8\cdot 2\cdot 4\cdot 2} \equiv 0 \pmod 2$, $\binom{k+5}{11} \equiv \frac{2\cdot 4\cdot 2\cdot(k-2)\cdot 2}{2\cdot 8\cdot 2\cdot 4\cdot 2} \equiv m \pmod 2$, $\binom{k+6}{13} \equiv \frac{(k+6)2\cdot 4\cdot 2\cdot(k-2)\cdot 2\cdot 4}{4\cdot 2\cdot 8\cdot 2\cdot 4\cdot 2} \equiv 0 \pmod 2$. Hence, by (*),

$$G_{k-1} \equiv \lambda + m\lambda^5 \pmod{(\lambda^7)}.$$

Hence, modulo the ideal (λ^8),

$$\begin{aligned}
&\overline{F}_k(F_k+\lambda G_{k-1})-\lambda^2 G_{k-1}\overline{F}_{k-1}\\
&=(1+\lambda+\lambda^2+\lambda^5+\lambda^6)(1+\lambda+\lambda^2+m\lambda^4+m\lambda^5+\lambda^2+m\lambda^6)\\
&\quad+\lambda^2(\lambda+m\lambda^5)(1+\lambda+\lambda^3+\lambda^4+\lambda^5)\\
&=(1+\lambda+\lambda^2+\lambda^5+\lambda^6)(1+\lambda+m\lambda^4+m\lambda^5+m\lambda^6)\\
&\quad+(\lambda^3+m\lambda^7)(1+\lambda+\lambda^3+\lambda^4+\lambda^5).
\end{aligned}$$

Thus $\alpha(H_n,4)=m+1$ and $\alpha(H_n,7)=(m+m+1)+(1+m)=m$. Hence either $\alpha(H_n,4)\neq 0$, or $\alpha(H_n,7)\neq 0$ and (6) is proved.

Finally, suppose $k\equiv -1 \pmod 8$. Write $k=m2^{s+1}-1$, with $s\geq 2$ and m odd. Recall that we are assuming $\operatorname{char}(F)=2$. We claim that $\alpha(H_n,2^s)=1$. Set

$$t=2^s.$$

Note that by 2.6 and 2.7, for $1\leq \ell\leq t$, $\alpha(F_k,\ell)=\binom{k+\ell}{2\ell}$,

$$\begin{aligned}
\alpha(G_{k-1},\ell)&=\binom{k+\ell}{2\ell+1},\\
\alpha(\overline{F}_k,t)&=\alpha(F_k,k-t)=\binom{2k-t}{t}=\binom{2k-2^s}{2^s},\\
\alpha(\overline{F}_k,t-1)&=\alpha(F_k,k-(t-1))=\binom{2k-(t-1)}{t-1}=\binom{2k-2^s+1}{2^s-1}\ \text{and}\\
\alpha(\overline{F}_{k-1},t-2)&=\alpha(F_{k-1},(k-1)-(t-2))=\binom{2(k-1)-(t-2)}{t-2}\\
&=\binom{2k-2^s}{2^s-2}.
\end{aligned}$$

Using 2.3, we see that

$$\begin{aligned}
&F_k\equiv 1+\lambda^t \pmod{(\lambda^{t+1})} \qquad G_{k-1}\equiv 1 \pmod{(\lambda^t)}\\
&\alpha(\overline{F}_k,t)=0,\quad \alpha(\overline{F}_k,t-1)=1,\quad \alpha(\overline{F}_{k-1},t-2)=1.
\end{aligned}$$

Hence $\alpha(H_n,t)$ is the coefficient of λ^t in the polynomial

$$(1+\lambda^{t-1})(1+\lambda+\lambda^t)+\lambda^2\lambda^{t-2}$$

which is 1.

2.14. *Suppose* $\operatorname{char}(\mathbb{F})=3$ *and* $n=2k$. *Let* $\beta\in\{1,-1\}$. *Set* $u=u_k^n(1)$, $h=uu^t$ *and*

$$\begin{aligned}
a(\beta)&=u_1^k(1)u_2^k(1)\cdots u_{k-2}^k(1)u_{k-1}^k(\beta)\\
b(\beta)&=u_1^k(-\beta)u_2^k(-1)u_3^k(-1)\cdots u_{k-1}^k(-1)\\
X(\beta)&=\operatorname{diag}(a(\beta),\{b(\beta)\}^{-1})u.
\end{aligned}$$

Then:

(1) $h = \operatorname{diag}(I_{k-1}, s, I_{k-1})$, *with* $s^{-1} = \begin{bmatrix} 2 & \bar{1} \\ \bar{1} & 1 \end{bmatrix}$.

(2) *If* $x = (a(\beta))^t a(-\beta)$ *and* $y = b(-\beta)(b(\beta))^t$, *then*

$$F_x = F_y = F_k - \lambda G_{k-2}.$$

(3) *Suppose* $k \equiv 1 \pmod 3$. *Set* $X = X(\beta)$, $Y = X(-\beta)$ *and* $L_n[\lambda] = F_{X^tY}$. *Then* $\alpha(L_n, 1) = -1$.

Proof. (1) is obvious. For (2), note that

$$a(\beta) = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 1 & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & 1 & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \beta & 1 \end{bmatrix} \text{ and } b(\beta) = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \bar{\beta} & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \bar{1} & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \bar{1} & 1 & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \bar{1} & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \bar{1} & 1 \end{bmatrix}.$$

Hence

$$x = \begin{bmatrix} 1 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & 1 & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 & \beta \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & 1 & 1 & 0 & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 1 & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \bar{\beta} & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & 2 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 2 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & 1 & 2 & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & 1 & 0 & \beta \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \bar{\beta} & 1 \end{bmatrix},$$

$$y = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \beta & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \bar{1} & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & \bar{1} & 1 & 0 & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & \bar{1} & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \bar{1} & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & \bar{\beta} & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & \bar{1} & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \bar{1} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & 1 & \bar{1} & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 & \bar{1} \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \bar{\beta} & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \beta & 0 & \bar{1} & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \bar{1} & 2 & \bar{1} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & \bar{1} & 2 & \bar{1} & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & \bar{1} & 2 & \bar{1} \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \bar{1} & 2 \end{bmatrix}.$$

To compute F_x expand $\det(\lambda I_k - x)$ along the last row. Thus

$$F_x = (\lambda - 1)F_{M_{k,k}(x)} + G_{k-2}$$

(since $\beta^2 = 1$). Also it is easy to see that

(i) $$F_{M_{k,k}(x)} = \lambda G_{k-2} - G_{k-3}.$$

Thus

(ii) $$F_x = (\lambda - 1)\{\lambda G_{k-2} - G_{k-3}\} + G_{k-2}.$$

Expanding F_y along the first row we see that $F_y = F_x$. Recall now from 2.6, that $F_k = (\lambda - 1)G_{k-1} - G_{k-2}$ and that $G_{k-1} = (\lambda - 2)G_{k-2} - G_{k-3} = (\lambda + 1)G_{k-2} - G_{k-3}$. Hence

(iii) $$F_k = (\lambda - 1)\{(\lambda + 1)G_{k-2} - G_{k-3}\} - G_{k-2}.$$

Thus, from (ii) and (iii) we see that $F_k - F_x = (\lambda - 1)G_{k-2} - 2G_{k-2} = (\lambda - 1)G_{k-2} + G_{k-2} = \lambda G_{k-2}$. This shows (2).

We proceed with the proof of (3). Note that $X^tY = u^t\mathrm{diag}\,((a(\beta))^t, (\{b(\beta)\}^{-1})^t)\mathrm{diag}\,(a(-\beta), \{b(-\beta)\}^{-1})u = u^t\mathrm{diag}\,(x, y^{-1})u$, with x and y as in (1). Now X^tY is conjugate to $h\mathrm{diag}\,(x, y^{-1})$, so we can use 2.9 to compute L_n. By 2.9 and (1),

(iv) $$L_n = \{F_x + \lambda F_{M_{k,k}(x)}\}F_{y^{-1}} - \lambda^2 F_{M_{k,k}(x)}F_{M_{1,1}(y^{-1})}.$$

Thus, by (iv),

(v) $$\alpha(L_n, 1) = \alpha(\{F_x + \lambda F_{M_{k,k}(x)}\}F_{y^{-1}}, 1).$$

Now, by (i) and (ii), $F_x + \lambda F_{M_{k,k}(x)} = F_k - \lambda G_{k-2} + \lambda\{\lambda G_{k-2} - G_{k-3}\}$. So

(vi) $$F_x + \lambda F_{M_{k,k}(x)} = F_k - \lambda G_{k-2} - \lambda G_{k-3} + \lambda^2 G_{k-2}.$$

Hence, by (v) and (vi),

(vii) $$\alpha(L_n, 1) = \alpha(\{F_k - \lambda G_{k-2} - \lambda G_{k-3}\}F_{y^{-1}}, 1).$$

Now modulo the ideal (λ^2), $F_k \equiv (-1)^k(1-\lambda)$, $\lambda G_{k-2} \equiv (-1)^{k-2}\binom{k-1}{1}\lambda \equiv 0$, $\lambda G_{k-3} \equiv (-1)^{k-3}\binom{k-2}{1}\lambda \equiv (-1)^{k-2}\lambda \equiv (-1)^k\lambda$. Thus

(viii) $$F_k - \lambda G_{k-2} - \lambda G_{k-3} \equiv (-1)^k(1+\lambda) \pmod{(\lambda^2)}.$$

Now, by (2), $F_y = F_k - \lambda G_{k-2} = (\lambda^k - \lambda^{k-1} + \cdots) - (\lambda^{k-1} + \cdots) = \lambda^k + \lambda^{k-1} + \cdots$. It follows from 2.7.1, that

(ix) $$F_{y^{-1}} \equiv (-1)^k(1+\lambda) \pmod{(\lambda^2)}.$$

Hence by (vii), (viii) and (ix), $\alpha(L_n, 1) = \alpha((1+\lambda)^2, 1) = -1$, and (3) is proved.

2.15. *Suppose $n = 2k$. Let $\alpha \in \mathbb{F}^*$ and set $u = u_k^n(\alpha)$. Let $X = \operatorname{diag}(a_k, b_k^{-1})u$ and set $x = a_k^t a_k^{-1}$ and $y = b_k^{-1}b_k^t$. Then*

$$\alpha\,(F_{X^tX^{-1}}, 1) = \alpha^2 - 2.$$

Proof. Note that $X^tX^{-1} = u^t\operatorname{diag}(a_k^t, (b_k^t)^{-1})u^{-1}\operatorname{diag}(a_k^{-1}, b_k)$. A moment of thought will convince the reader that u commutes with $\operatorname{diag}(a_k^t, (b_k^t)^{-1})$, hence

(i) $$X^tX^{-1} = u^tu^{-1}\operatorname{diag}(x, y^{-1}).$$

Set $h = u^tu^{-1}$ and $g = \operatorname{diag}(x, y^{-1})$. Then

(ii) $$h^{-1} = \operatorname{diag}\left(I_{k-1}, \begin{bmatrix} 1 & \bar{\alpha} \\ \alpha & 1-\alpha^2 \end{bmatrix}, I_{k-1}\right).$$

We use 2.9, with $A = x$, $B = y^{-1}$, $h = u^tu^{-1}$. By (i), $X^tX^{-1} = hg$. By 2.9,

(iii) $$F_{X^tX^{-1}} = F_xF_{y^{-1}} + (\beta_{22}-1)\lambda F_xF_{M_{1,1}(y^{-1})} + (\beta_{11}-1)\lambda F_{M_{k,k}(x)}F_{y^{-1}} + \det\begin{bmatrix} (\beta_{11}-1)\lambda & \beta_{12}\lambda \\ \beta_{21}\lambda & (\beta_{22}-1)\lambda \end{bmatrix} F_{M_{k,k}(x)}F_{M_{1,1}(y^{-1})}.$$

Of course, by (ii), here $\beta_{11} = 1$, $\beta_{12} = -\alpha$, $\beta_{21} = \alpha$ and $\beta_{22} = 1-\alpha^2$. Note that by 2.5,

(iv) $$F_x = F_{y^{-1}} = Q_k.$$

Further, by 2.11.1 and 2.11.4, $F_{M_{1,1}(y^{-1})} = F_{(b_{k-1}^{-1})b_{k-1}}$, so by 2.5,

(v) $$F_{M_{1,1}(y^{-1})} = Q_{k-1}.$$

Now by (iii), (iv) and (v), we get

$$F_{X^tX^{-1}} = Q_k\{Q_k - \alpha^2\lambda Q_{k-1}\} + \alpha^2\lambda^2 \cdot F_{M_{k,k}(x)} \cdot F_{M_{1,1}(y^{-1})}.$$

Whence,

$$\begin{aligned}\alpha\,(F_{X^tX^{-1}},1) &= \alpha(Q_k\{Q_k - \alpha^2\lambda Q_{k-1}\},1)\\ &= (-1)^k\{(-1)^{k+1} - \alpha^2(-1)^{k-1}\} + (-1)^k(-1)^{k+1}\\ &= -1 + \alpha^2 - 1\\ &= \alpha^2 - 2.\end{aligned}$$

2.16. *Suppose* $\operatorname{char}(\mathbb{F}) = 3$*,* $n = 2k \geq 8$ *and that* $k \equiv 1 \pmod 3$*. Let* $\beta \in \{1,-1\}$ *and let* $a(\beta), b(\beta), X, Y$ *and* u *be as in* 2.14*. Set* $x = (a(\beta))^t(a(-\beta))^{-1}$ *and* $y = (b(-\beta))^{-1}(b(\beta))^t$*. Then*

$$F_x = F_y = \lambda^k + (-1)^k = F_{y^{-1}}, \tag{1}$$

$$\alpha\,(F_{X^tY^{-1}},1) = 1. \tag{2}$$

Proof. Note that $X^tY^{-1} = u^t\operatorname{diag}((a(\beta))^t, ((b(\beta))^t)^{-1})u^{-1}\operatorname{diag}((a(-\beta))^{-1}, b(-\beta))$. Now a moment of thought will convince the reader that u commutes with $\operatorname{diag}((a(\beta))^t, ((b(\beta))^t)^{-1})$, hence

$$X^tY^{-1} = u^tu^{-1}\operatorname{diag}(x, y^{-1}). \tag{i}$$

Set $h = u^tu^{-1}$ and $g = \operatorname{diag}(x, y^{-1})$. Then

$$h^{-1} = \operatorname{diag}\left(I_{k-1}, \begin{bmatrix}1 & \bar{1}\\ 1 & 0\end{bmatrix}, I_{k-1}\right). \tag{ii}$$

Next note that, by 1.11, $\operatorname{diag}(a(\beta), \{b(\beta)\}^{-1}), \operatorname{diag}(a(-\beta), \{b(-\beta)\}^{-1}) \in \operatorname{Fix}(\tau)$, so, by 2.4, $F_x = F_y$. Also if $F_y = \lambda^k + (-1)^k$, then, by 2.7.1, $F_{y^{-1}} = \lambda^k + (-1)^k$. We now use 2.12.3 to compute F_y. Take in 2.12.3, $B_1 = b_k(-\beta, 1, \dots, 1)$ and $C_1 = b_k(\beta, 1, \dots, 1)$ (notice that $\beta_1 = -\beta$ and $\gamma_1 = \beta$). By 2.12.3, $F_y = (\lambda - 1)Q_{k-1} - \beta^2\lambda Q_{k-2}$ and since $\beta^2 = 1$, $F_y = (\lambda-1)Q_{k-1} - \lambda Q_{k-2}$. Notice now that $\lambda Q_{k-1} = \lambda^k - Q_{k-1} + (-1)^{k-1}$, and $\lambda Q_{k-2} = Q_{k-1} - (-1)^{k-1}$. Hence $F_y = (\lambda^k - Q_{k-1} + (-1)^{k-1}) - Q_{k-1} - (Q_{k-1} - (-1)^{k-1}) = \lambda^k - 3Q_{k-1} + 2(-1)^{k-1}$. Since $\operatorname{char}(\mathbb{F}) = 3$, (1) follows.

Next, $y^{-1} = (\{b(\beta)\}^{-1})^t(b(-\beta))$. By 2.11.4 and 2.11.1, $M_{1,1}(y^{-1}) = (b_{k-1}^{-1})^tb_{k-1}$ and so $F_{M_{1,1}(y^{-1})} = F_{(b_{k-1}^{-1})^tb_{k-1}}$, hence by 2.5

$$F_{M_{1,1}(y^{-1})} = Q_{k-1}. \tag{iii}$$

For (2), we use 2.9, with $A = x$, $B = y^{-1}$, $g = \operatorname{diag}(A, B)$ and $h = u^tu^{-1}$. By 2.9,

$$\begin{aligned}F_{hg} = F_AF_B &+ (\beta_{22} - 1)\lambda F_AF_{M_{1,1}(B)} + (\beta_{11} - 1)\lambda F_{M_{k,k}(A)}F_B\\ &+ \det\begin{bmatrix}(\beta_{11} - 1)\lambda & \beta_{12}\lambda\\ \beta_{21}\lambda & (\beta_{22} - 1)\lambda\end{bmatrix} F_{M_{k,k}(A)}F_{M_{1,1}(B)}.\end{aligned}$$

By (i), $X^tY^{-1} = hg$ and by (ii), here $\beta_{11} = 1$, $\beta_{12} = -1$, $\beta_{21} = 1$ and $\beta_{22} = 0$. Using 2.9, (1) and (iii), we get

$$F_{X^tY^{-1}} = (\lambda^k + (-1)^k)\{\lambda^k + (-1)^k - \lambda Q_{k-1}\} + \lambda^2 F_{M_{k,k}(A)} \cdot F_{M_{1,1}(B)}.$$

Hence, $\alpha(F_{X^tY^{-1}}, 1) = \alpha((\lambda^k + (-1)^k)\{\lambda^k + (-1)^k - \lambda Q_{k-1}\}, 1) = 1$, as is easily checked.

3. The Special Linear Groups.

In this section we prove Theorem 1.6 for the groups $L_n(q)$. We let $L = SL_n(\mathbb{F})$. Of course all notation and definitions introduced in Section 1 are maintained here. By 1.7 and 1.9.2, all we have to do is to find an element $X \in L$, such that $B(X, X^t)$. We take

$$X = a_n.$$

3.1. *Let $S \in \{X^tX, X^tX^{-1}, X^t\}$ and let $R \in \Delta^{\leq 2}(X) \cap \Delta^{\leq 1}(S)$. Then v_1 is a characteristic vector of R.*

Proof. Let $h \in \Delta^{\leq 1}(X) \cap \Delta^{\leq 1}(R)$. Note that since X is unipotent and $[X, h] \in Z(L)$, $[X, h] = 1$. By 1.13, there exists $\beta \in \mathbb{F}$ and $1 \leq r < n$, such that $h - \beta I_n \in \mathcal{T}_n(r)$ (see notation in 1.1.10). Put $T = h - \beta I_n$, $j = m = r$ and $\ell = 0$. We'll show that S, T, R, j, m and ℓ satisfy the hypotheses of 1.15. Hence, by 1.15, v_1 is a characteristic vector of R.

Since $(X^t)_{i,i+1} = 1$, while, $(X^t)_{i,k} = 0$, for all $1 \leq i \leq n-1$ and all $i + 1 < k \leq n$, and since X^ϵ is unipotent lower triangular, for $\epsilon \in \{1, -1\}$, it is easy to see that hypothesis (a) of 1.15 is satisfied. Of course $\mathcal{V}_j = \mathcal{V}_r \subseteq \ker(T)$. By definition, $v_{j+1} \notin \ker(T)$. Since $\mathcal{V}_m = \mathcal{V}_r = \ker(T)$ and since R centralizes $T, \mathcal{V}_m$ is R-invariant. By now we verified all hypotheses of 1.15 and the proof of 3.1 is complete.

3.2. *Let $S = XX^t$. Then:*

(1) *If* $\mathrm{char}\,(\mathbb{F}) \neq 3$, *or* $n - 2 \not\equiv 0 \pmod 3$, *then either* $\alpha(F_S, n-1) \neq 0$ *or* $\alpha(F_S, 1) \neq 0$.
(2) *If* $\mathrm{char}\,(\mathbb{F}) = 3$ *and* $n - 2 \equiv 3, 6 \pmod 9$, *then* $\alpha(F_S, n-2) \neq 0 \neq \alpha(F_S, n-3)$.
(3) *If* $\mathrm{char}\,(\mathbb{F}) = 3$ *and* $n - 2 \equiv 0 \pmod 9$, *then* $\alpha(F_S, n-2) \neq 0 \neq \alpha(F_S, n-5)$.

Proof. By definition 1.2.4, $F_S = F_n$. So by 2.6.4,

$$F_S = \sum_{\ell=0}^{n} (-1)^{n+\ell} \binom{n+\ell}{2\ell} \lambda^\ell.$$

In particular, $\alpha(F_S, n-1) = 1 - 2n$ and $\alpha(F_S, 1) = (-1)^{n+1}\binom{n+1}{2}$. Let $p = \mathrm{char}\,(\mathbb{F})$ and suppose $\alpha(F_S, n-1) = \alpha(F_S, 1) = 0$. It is easy to check that we must have $p = 3$ and $n \equiv -1 \pmod 3$. So suppose $\mathrm{char}\,(\mathbb{F}) = 3$ and

$n \equiv -1 \pmod 3$. Note that $\alpha(F_S, n-2) = (n-1)(2n-3)$, so $\alpha(F_S, n-2) \neq 0$. If $n-2 \equiv 3, 6 \pmod 9$, then $\alpha(F_S, n-3) = -\binom{2n-3}{3} \not\equiv 0 \pmod 3$. Finally, if $n-2 \equiv 0 \pmod 9$, then $\alpha(F_S, n-5) = -\binom{2n-5}{5} \not\equiv 0 \pmod 3$. We remark that when $n = 2$, Δ is disconnected and there exists no path from X to S in Δ, so evidently $B(X, X^t)$ holds.

3.3. (1) *Let* $S \in \{X^t X, X^t X^{-1}, X^t\}$, *then* $d(X, S) > 3$.
(2) $\Delta(L)$ *is balanced.*

Proof. Let $R \in \Delta^{\leq 2}(X) \cap \Delta^{\leq 1}(S)$. By 3.1,

(i) $$v_1 \text{ is a characteristic vector of } R.$$

Note that for all $1 \leq i \leq n-1$, $v_i S = u + v_{i+1}$, with $u \in \mathcal{V}_i$. Thus

(ii) $$\langle \mathcal{O}(v_1, S) \rangle = V.$$

Now if $S = X^t$, then, by (i), (ii) and 1.14.1, $R \in Z(GL(V))$, a contradiction.

Suppose $S = X^t X$. Note that by 3.2, $\gcd\Big\{\{i : \alpha_{n-i} \neq 0\} \cup \{n\}\Big\} = 1$, thus, by (i), (ii) and 1.14.5, $R \in Z(GL(V))$, a contradiction. Finally suppose $S = X^t X^{-1}$. Then, by 2.5, $\alpha(F_S, n-1) \neq 0$, and again, by 1.14.5, $R \in Z(GL(V))$, a contradiction. This shows (1). (2) follows immediately from (1), since, by definition, $B(X, X^t)$ and then, by 1.9.2, $B(X^t, X)$, so by definition, $\Delta(L)$ is balanced.

4. The Symplectic Groups and Unitary Groups in even dimension.

In this section $n = 2k \geq 4$. Further, $\mathbb{F}$ is a field of order q^2 and $\mathbb{K} \leq \mathbb{F}$ is a field of order q. L is one of the following groups. Either $L = \mathrm{Fix}\,(\tau)$, where $\tau : SL_n(\mathbb{K}) \to SL_n(\mathbb{K})$ is the automorphism defined in 1.4.4, or $L = \mathrm{Fix}\,(\tau\sigma_q)$, where $\tau\sigma_q : SL_n(\mathbb{F}) \to SL_n(\mathbb{F})$ is the automorphism defined in 1.4.4 and 1.4.5. Thus, by 1.12.3, in the first case $L \simeq Sp_n(q)$, and in the second case $L \simeq SU_n(q)$. The purpose of this section is to prove that Theorem 1.6 holds for (the simple version of) L. We'll pick two elements $X, Y \in L$ and show that $B(X, Y^t)$ and $B(Y, X^t)$. By 1.9.1, also $B(Y^t, X)$ and thus the elements X, Y show that $\Delta(L)$ is balanced. In most cases, we'll take $X = Y$, but when $\mathrm{char}\,(\mathbb{F}) = 3$, it turns out that we must pick $Y \neq X$. For the moment we fix elements $\beta_1, \dots, \beta_{k-1}, \gamma_1, \dots, \gamma_{k-1}$, $\alpha \in \mathbb{K}^*$. Using the notation in 1.1.8 we let

$$a = a_k(\beta_1, \dots, \beta_{k-1}) \qquad a_1 = a_k(\gamma_1, \dots, \gamma_{k-1})$$
$$b = b_k(\beta_1, \dots, \beta_{k-1}) \qquad b_1 = b_k(\gamma_1, \dots, \gamma_{k-1})$$
$$g = \mathrm{diag}\,(a, b^{-1}) \qquad g_1 = \mathrm{diag}\,(a_1, b_1^{-1})$$
$$u = u_k^n(\alpha)$$
$$X = gu \qquad Y = g_1 u.$$

Towards the end of Section 4 we'll specialize and give concrete values to β_i, γ_i and α. Note that by 1.11, $X, Y \in L$.

4.1. *Let $u = u_k^n(\alpha)$. Then:*

$$uu^t = \operatorname{diag}\left(I_{k-1}, \begin{bmatrix}1 & \alpha \\ \alpha & \alpha^2+1\end{bmatrix}, I_{k-1}\right) \tag{1}$$
$$(uu^t)^{-1} = \operatorname{diag}\left(I_{k-1}, \begin{bmatrix}\alpha^2+1 & \overline{\alpha} \\ \overline{\alpha} & 1\end{bmatrix}, I_{k-1}\right).$$

$$u^{-1}u^t = \operatorname{diag}\left(I_{k-1}, \begin{bmatrix}1 & \alpha \\ \overline{\alpha} & 1-\alpha^2\end{bmatrix}, I_{k-1}\right) \tag{2}$$
$$(u^{-1}u^t)^{-1} = \operatorname{diag}\left(I_{k-1}, \begin{bmatrix}1-\alpha^2 & \overline{\alpha} \\ \alpha & 1\end{bmatrix}, I_{k-1}\right).$$

(3) $[u, g^t] = 1$.

Proof. This is obvious.

4.2. *Let $\epsilon \in \{1, -1\}$. Then:*

(1) $XY^t = guu^tg_1^t$, $(X, Y^t)^{-1} = (g_1^t)^{-1}(uu^t)^{-1}g^{-1}$.
(2) $X^{-1}Y^t = u^{-1}u^tg^{-1}g_1^t$ *and* $(X^{-1}Y^t)^{-1} = (g_1^t)^{-1}g(u^{-1}u^t)^{-1}$.
(3) $X = \begin{bmatrix}a & 0_{k,k} \\ E & b^{-1}\end{bmatrix}$ *with E some $k \times k$ matrix, such that $E_{1,k} = \alpha$.*
(4) $X^\epsilon Y^t = \begin{bmatrix}a^\epsilon a_1^t & R_{1,2} \\ R_{2,1} & R_{2,2}\end{bmatrix}$ $(X^\epsilon Y^t)^{-1} = \begin{bmatrix}R'_{1,1} & R'_{1,2} \\ R'_{2,1} & b_1^tb^\epsilon\end{bmatrix}$ *with $R_{i,j}$ and $R'_{i,j}$ some $k \times k$ matrices. Further, the first $k-1$ rows of $R_{1,2}$ are zero.*
(5) *Let $S \in \{Y^t, X^\epsilon Y^t\}$. Then for $1 \le i \le k-1$, $v_iS = v + \delta_{i+1}v_{i+1}$, with $v \in \mathcal{V}_i$ and $\delta_{i+1} \in \mathbb{K}^*$.*
(6) *Let $S \in \{Y^t, X^\epsilon Y^t\}$. Then for $k \le i \le n-1$, $v_iS^{-1} = v + \delta_{i+1}v_{i+1}$, with $v \in \mathcal{V}_i$ and $\delta_{i+1} \in \mathbb{K}^*$.*
(7) *Let $S \in \{Y^t, X^\epsilon Y^t\}$, then $V = \langle \mathcal{O}(v_1, S)\rangle$.*
(8) *Let $S \in \{Y^t, X^\epsilon Y^t\}$, then $S_{k,n} \neq 0$.*

Proof. (1) is obvious. For (2), we have $X^{-1}Y^t = u^{-1}g^{-1}u^tg_1^t$. By 4.1.3, $[g^{-1}, u^t] = 1$, and (2) follows. (3) is clear, the $(1,k)$-entry of E is $\alpha\cdot(b^{-1})_{1,1} = \alpha$.

To show(4) and (5), let $1 \le i \le k-1$, then $v_iu^{-1}u^t = v_i$, so $v_iX^{-1}Y^t = v_ig^{-1}g_1^t$. Also $v_ig \in \mathcal{V}_i$, so $v_ig(uu^t) = v_ig$ and $v_iXY^t = v_igg_1^t$. We conclude that:

$$\text{For } 1 \le i \le k-1,\ v_iX^\epsilon Y^t = v_ig^\epsilon g_1^t. \tag{i}$$

Now the shape of $X^\epsilon Y^t$ follows from (3) and (i), since, by (i), the first $k-1$ rows of $R_{1,2}$ are zero. Also the shape of $(X^\epsilon Y^t)^{-1}$, follows from (3). For (5), we use (i). Note that a^ϵ is unipotent, lower triangular and a_1^t is upper triangular unipotent with $(a_1^t)_{i,j} = 0$, for $j > i+1$, and $(a_1^t)_{i,i+1} \neq 0$. This easily implies (5), for $S = X^\epsilon Y^t$. For $S = Y^t$, $v_i Y^t = v_i + \beta_{k-i} v_{i+1}$, for all $1 \leq i \leq k-1$, thus (5) holds for Y^t as well.

For (6), note that for $h \in \{b_1^t, b_1^t b^\epsilon\}, h_{i,j} = 0$, for $j > i+1$, and $h_{i,i+1} \neq 0$, for all $1 \leq i \leq k-1$. This clearly holds for b_1^t and since this holds for b_1^t and b^ϵ is unipotent lower triangular, it also hold for $b_1^t b^\epsilon$. Thus, by (4), (6) holds for $S \in \{Y^t, X^\epsilon Y^t\}$ and $k+1 \leq i \leq n-1$. We compute that $v_k(Y^t)^{-1} = v_k(g_1^t)^{-1}(u^t)^{-1} = v_k(u^t)^{-1} = v_k - \alpha v_{k+1}$. Also $v_k(X^\epsilon Y^t)^{-1} = v_k(Y^t)^{-1}X^{-\epsilon} = (v_k - \alpha v_{k+1})X^{-\epsilon} = v_k X^{-\epsilon} - \alpha v_{k+1} X^{-\epsilon}$. Now $v_k X^{-\epsilon} \in \mathcal{V}_k$, and $v_{k+1} X^{-\epsilon} \equiv v_{k+1} \pmod{\mathcal{V}_k}$, so (6) follows. (7) follows from (5) and (6), since by (5), $\mathcal{V}_k \subseteq \langle \mathcal{O}(v_1, S)\rangle$, and then by (6), $\langle \mathcal{O}(v_1, S)\rangle = V$.

Finally, to show (8), note that $v_k X^\epsilon = v + v_k$, with $v \in \mathcal{V}_{k-1}$, and by (5), $vY^t \in \mathcal{V}_k$. Thus for $S \in \{Y^t, X^\epsilon Y^t\}$, $S_{k,n} = (Y^t)_{k,n}$. Now $v_k Y^t = v_k u^t g_1^t = (v_k + \alpha v_{k+1}) g_1^t = v_k + \alpha v_{k+1} g_1^t$. Now it is easy to check that $(b_1^{-1})_{k,1} = \prod_{i=1}^k \gamma_i \neq 0$, thus $(g_1^t)_{k+1,n} = (b_1^{-1})_{k,1} \neq 0$, hence $(Y^t)_{k,n} = (g_1^t)_{k+1,n} \neq 0$ and (8) is proved.

4.3. *Let $\epsilon \in \{1, -1\}$ and let $S \in \{Y^t, X^\epsilon Y^t\}$. Let $R \in \Delta^{\leq 2}(X) \cap \Delta^{\leq 1}(S)$. Then v_1 is a characteristic vector of R.*

Proof. Let $h \in \Delta^{\leq 1}(X) \cap \Delta^{\leq 1}(R)$. Then, $[h, X] = 1$, so by 4.2.3 and 1.13, there exists $0 \neq \beta \in \mathbb{K}$, and $1 \leq r \leq n-1$, such that $h - \beta I_n \in \mathcal{T}_n(r)$. We use 1.15. We take in 1.15, $T = h - \beta I_n$. Note that R commutes with h and hence with T.

Suppose first that $r \leq k-1$, we take in 1.15 $j = r = m$ and $\ell = 0$. Notice that by 4.2.5, hypothesis (a) of 1.15 is satisfied, hypothesis (b) and (c) of 1.15 are satisfied by definition, and we observed that hypothesis (e) of 1.15 is satisfied. Finally, since R centralizes T, $\mathcal{V}_r$ is R-invariant. Hence 1.15 completes the proof in this case.

Suppose next that $r \geq k$, we take in 1.15, $j = k-1$, $\ell = 1$ and $m = \dim(\mathrm{im}(T))$. Notice that $\mathrm{im}\,(T) = \mathcal{V}_m$ and $\mathrm{im}\,(T)$ is R-invariant. Also, by 4.2.8, $S_{k,n} \neq 0$, so clearly $v_k \notin \ker(ST)$ and hypothesis (c) of 1.15 holds. Thus 1.15 completes the proof in this case too.

From this point to the end of Section 4 we specialize and set:

$$\text{If } |\mathbb{K}| = 2, \text{ or } |\mathbb{K}| > 3, \text{ or } k \not\equiv 1 \pmod 3,$$
$$\beta_i = \gamma_i = 1, \text{ for all } 1 \leq i \leq k-1, \text{ in particular, } X = Y.$$
$$\text{If } |\mathbb{K}| = 3 \text{ and } k \equiv 1 \pmod 3,$$
$$\beta_i = \gamma_i = 1, \text{ for all } 2 \leq i \leq k-1 \text{ and } \beta_1 = -\gamma_1 = \beta.$$

4.4. (1) *If $|\mathbb{K}| > 3$, or $k \not\equiv 1 \pmod 3$, we can find $\alpha \in \mathbb{K}^*$ such that $\alpha(F_S, 1) \neq 0$ for all $S \in \{XY^t, X^{-1}Y^t\}$.*

(2) *If $|\mathbb{K}| = 3$, or $k \equiv 1 \pmod 3$, then for $\alpha = 1$ we have $\alpha(F_S, 1) \neq 0$, for all $S \in \{XY^t, X^{-1}Y^t\}$.*

(3) *If* $|\mathbb{K}| = 2$*, then for* $\alpha = 1$*,* $\gcd\Big\{\{i : \alpha(F_S, n-i) \neq 0\} \cup \{n\}\Big\}$ *is relatively prime to* 3*, for all* $S \in \{XY^t, X^{-1}Y^t\}$*.*

Proof. For (1), note that by our choice of X and Y, $X = Y$. Further, F_{XX^t} is the polynomial H_n of 2.13. Thus, $\alpha(F_{XX^t}, 1) = -\binom{k+1}{2} - (\alpha^2+2)k + 1$ by 2.13.2. Also, by 2.15, $\alpha(F_{X^{-1}X^t}, 1) = \alpha^2 - 2$. The reader may now easily verify (using also 2.13.3) that we can choose $\alpha \in \mathbb{K}^*$ as asserted in (1).

So suppose $|\mathbb{K}| = 3$ and $k \equiv 1 \pmod 3$. Then by 2.14.3, and 2.16, (2) holds. Finally assume $|\mathbb{K}| = 2$. Then (3) holds by 2.13.4-2.13.7 and by 2.15.

We now specialize further and choose α as in 4.4, in the respective cases.

4.5. *Set* $\Lambda = \Delta(L)$ *and let* $\epsilon \in \{1, -1\}$ *and let* $S \in \{Y^t, X^\epsilon Y^t\}$*. Then:*

(1) $d_\Lambda(X, S) > 3$.
(2) $B_\Lambda(X, Y^t)$ *and* $B_\Lambda(Y, X^t)$*.*
(3) $\Delta(L)$ *is balanced.*

Proof. Suppose $d_\Lambda(X, S) \leq 3$ and let $R \in \Lambda^{\leq 2}(X) \cap \Lambda^{\leq 1}(S)$. Of course $R \in \Delta^{\leq 2}(X) \cap \Delta^{\leq 1}(S)$, so by 4.3,

(i) $$v_1 \text{ is a characteristic vector of } R.$$

If $S = Y^t$, then $[R, S] = 1$, so by (i), 4.2.7 and 1.14.1, $R \in Z(L)$, a contradiction. So (1) holds in case $S = Y^t$. So assume $S = X^\epsilon Y^t$.

Suppose first that $|\mathbb{K}| > 3$, or $|\mathbb{K}| = 3$ and $k \not\equiv 1 \pmod 3$, then using 4.4.1, (i), 4.2.7 and 1.14.5, we see that $R \in Z(L)$, a contradiction. This shows (1) in this case. By (1), $B_\Lambda(X, Y^t)$ holds here, and since here $X = Y$, 1.9.2 implies (2) in this case.

Suppose $|\mathbb{K}| = 3$ and $k \equiv 1 \pmod 3$. Then using 4.4.2, (i), 4.2.7 and 1.14.5, we see that $R \in Z(L)$, a contradiction. Hence (1) holds here and by (1) and definition, $B_\Lambda(X, Y^t)$ holds in this case. By Symmetry $d(Y, X^t) > 3$ and $d(Y, Y^\epsilon X^t) > 3$. Thus $B_\Lambda(Y, X^t)$ also holds and (2) holds in this case as well.

Finally, suppose $|\mathbb{K}| = 2$. If $L \simeq Sp_n(q)$, then $Z(L) = 1$, so $[R, S] = 1$, and hence, by (i), 4.2.7 and 1.14.1, $R = 1$, a contradiction. So assume $L \simeq SU_n(q)$. Then $|\mathbb{F}^*| = 3$. Now 4.4.3, (i), 4.2.7 and 1.14.5 show that $R \in Z(L)$, a contradiction. Again we see that (1) holds, and since $X = Y$ here, (2) holds here (as above). Note that (2) implies (3) by 1.9 and by definition.

5. The Unitary and Orthogonal Groups in odd dimension.

In this section $\mathbb{F}$ is a field of order q^2 and $\mathbb{K} \leq \mathbb{F}$ is the subfield of order q. We let $n = 2k+1 \geq 3$ be an odd integer and $U \simeq SU(n, \mathbb{F}) \leq SL(n, \mathbb{F})$ is the special unitary group. We view U as the fixed points of the automorphism

$\tau\sigma_q : SL(n,\mathbb{F}) \to SL(n,\mathbb{F})$, described in 1.12.3. We denote by $U \geq O \simeq SO(n,\mathbb{K})$, the subgroup $O = U \cap SL(n,\mathbb{K})$. L denotes one of the groups U or O. When $L = O$, we assume that $n \geq 7$ and that q is odd (this is because if q is even or $n < 7$, O' is either not simple, or isomorphic to simple groups that we handled earlier). We continue the notation of Section 1. In particular, V is a vector space of dimension n over $\mathbb{F}$.

Throughout this section $\Lambda = \Delta(L)$. The purpose of this section is to prove that when $L'/Z(L')$ is simple, $\Delta(L')$ is balanced (and hence, by 1.7, $\Delta(L'/Z(L'))$ is balanced). For that we'll indicate elements $X, Y \in L'$ such that $B_\Lambda(X, Y^t)$ and $B_\Lambda(Y, X^t)$ (see 1.10).

Notation 5.1. (1) given an element $r = \operatorname{diag}(I_{k-1}, s, I_{k-1}) \in GL(n,\mathbb{F})$, we denote $s(r) := s$ (note that $s \in GL_3(\mathbb{F})$).

(2) Let $\theta \in \mathbb{F}^*$. We denote by $u_0(\theta) = \operatorname{diag}(I_{k-1}, s, I_{k-1})$, with

$$s = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \theta & 0 & 1 \end{bmatrix}.$$

(3) Whenever we write $u_i(\alpha)$, we mean $u_i^n(\alpha)$ (see 1.1.7).

5.2. *Let $\alpha \in \mathbb{F}^*$ and $\beta_1, \dots, \beta_{k-1} \in \mathbb{K}^*$. Set $a = a_k(\beta_1, \dots, \beta_{k-1})$, $b = b_k(\beta_1, \dots, \beta_{k-1})$, $B = b_{k+1}(\alpha, \beta_1, \dots, \beta_{k-1})$ and $g = \operatorname{diag}(a, 1, b^{-1})$. Let $u = u_k(\alpha)u_{k+1}(\alpha^q)u_0(\theta)$. Then:*

(1) $g \in O$.

(2) $gu_{k+1}^n(\alpha) = \operatorname{diag}(a, B^{-1})$.

(3) $[g, u^t] = 1$.

Proof. (1) is 1.11. For (2), note that $g = \operatorname{diag}(a, z)$, with

$$z = u_k^{k+1}(\beta_{k-1})u_{k-1}^{k+1}(\beta_{k-2}) \cdots u_2^{k+1}(\beta_1).$$

Also $u_{k+1}(\alpha) = \operatorname{diag}(I_k, u_1^{k+1}(\alpha))$. Thus $gu_{k+1}(\alpha) = \operatorname{diag}(a, h)$, with $h = zu_1^{k+1}(\alpha) = u_k^{k+1}(\beta_{k-1})u_{k-1}^{k+1}(\beta_{k-2}) \cdots u_2^{k+1}(\beta_1)u_1^{k+1}(\alpha) = B^{-1}$.

(3) follows from the fact that $(u_k(\alpha))^t$, $(u_{k+1}(\alpha^q))^t$, and $(u_0(\theta))^t$ commute with g.

5.3. *Let $\alpha, \beta, \theta \in \mathbb{F}$ and set $u = u_k(\alpha)u_{k+1}(\beta)u_0(\theta)$. Then:*

(1) $s(u_k(\alpha)u_{k+1}(\beta)) = \begin{bmatrix} 1 & 0 & 0 \\ \alpha & 1 & 0 \\ 0 & \beta & 1 \end{bmatrix}$.

(2) $s(u_{k+1}(\beta)u_k(\alpha)) = \begin{bmatrix} 1 & 0 & 0 \\ \alpha & 1 & 0 \\ \alpha\beta & \beta & 1 \end{bmatrix}$.

(3)

$$\begin{aligned} u_0(\theta) &= u_k(1)u_{k+1}(-\theta)u_k(-1)u_{k+1}(\theta)\\ &= u_{k+1}(1)u_k(\theta)u_{k+1}(-1)u_k(-\theta). \end{aligned}$$

(4) $u_0(\theta)\tau = u_0(-\theta)$.
(5) $u \in \mathrm{Fix}\,(\tau\sigma_q)$ *iff* $\beta = \alpha^q$ *and* $\theta + \theta^q = \alpha^{q+1}$.

Proof. (1) and (2) are easy to check. For (3) we have

$$\begin{aligned} &s\Big\{u_k(1)u_{k+1}(-\theta)u_k(-1)u_{k+1}(\theta)\Big\}\\ &= \begin{bmatrix} 1 & 0 & 0\\ 1 & 1 & 0\\ 0 & \bar\theta & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0\\ \bar 1 & 1 & 0\\ 0 & \theta & 1 \end{bmatrix}\\ &= \begin{bmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ \theta & 0 & 1 \end{bmatrix} = s(u_0(\theta)) \end{aligned}$$

and

$$\begin{aligned} &s\Big\{u_{k+1}(1)u_k(\theta)u_{k+1}(-1)u_k(-\theta)\Big\}\\ &= \begin{bmatrix} 1 & 0 & 0\\ \theta & 1 & 0\\ \theta & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0\\ \bar\theta & 1 & 0\\ \theta & \bar 1 & 1 \end{bmatrix}\\ &= \begin{bmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ \theta & 0 & 1 \end{bmatrix} = s(u_0(\theta)). \end{aligned}$$

For (4), note that by (3), $u_0(\theta)\tau = \{u_k(1)u_{k+1}(-\theta)u_k(-1)u_{k+1}(\theta)\}\tau = u_{k+1}(1)u_k(-\theta)u_{k+1}(-1)u_k(\theta) = u_0(-\theta)$.

For (5), we have

$$s(u) = \begin{bmatrix} 1 & 0 & 0\\ \alpha & 1 & 0\\ \theta & \beta & 1 \end{bmatrix}.$$

Now, by (4), $u\tau\sigma_q = u_{k+1}(\alpha^q)u_k(\beta^q)u_0(-\theta^q)$, so

$$s(u\tau\sigma_q) = \begin{bmatrix} 1 & 0 & 0\\ \beta^q & 1 & 0\\ (\alpha\beta - \theta)^q & \alpha^q & 1 \end{bmatrix}.$$

So the lemma follows.

Notation 5.4. Let $\alpha, \theta \in \mathbb{F}$ such that $\theta + \theta^q = \alpha^{q+1}$.

(1) We denote

$$u^n(\alpha,\theta) = u(\alpha,\theta) = u_k(\alpha)u_{k+1}(\alpha^q)u_0(\theta)$$
$$= u_{k+1}(\alpha^q)u_k(\alpha)u_0(-\theta^q).$$

(2) We denote $X(\alpha,\theta) = \text{diag}\,(a_k, 1, b_k^{-1})u(\alpha,\theta)$.

Note that we denote $u(\alpha,\theta)$ and $X(\alpha,\theta)$ only when $\theta+\theta^q = \alpha^{q+1}$, so that $u(\alpha,\theta)$, $X(\alpha,\theta) \in U$.

5.5. *Let $\alpha,\beta \in \mathbb{F}^*$ and let $u = u_1^{k+1}(-\alpha)$, $w = u_1^{k+1}(-\beta)$ and $\epsilon \in \{1,-1\}$. Then:*

(1) $(w^t u^\epsilon)^{-1} = \text{diag}\left(\begin{bmatrix} 1 & \beta \\ \epsilon\alpha & \epsilon\alpha\beta+1 \end{bmatrix}, I_{k-1}\right)$.

(2) $w^t u^\epsilon = \text{diag}\left(\begin{bmatrix} 1+\epsilon\alpha\beta & \beta \\ -\epsilon\alpha & 1 \end{bmatrix}, I_{k-1}\right)$.

Proof. This is obvious.

5.6. *Suppose* char $(\mathbb{F}) = 3$. *Then:*

(1) *For $B = b_{k+1}$, $F_{B^tB} = F_{k+1}$ and $F_{B^tB^{-1}} = Q_{k+1}$, in particular $F_{B^tB}[-1] \neq 0$ and $F_{B^tB^{-1}}[-1] = Q_{k+1}[-1] = (-1)^{k+1}(k+2)$.*

(2) *Suppose $k \geq 4$ and let $B = b_{k+1}(1,1,1,\beta_4,1,\dots,1)$ and $C = b_{k+1}(1,1,1,\gamma_4,1,\dots,1)$, with $\beta_4\gamma_4 = -1$. Then for $\{T,Z\} = \{B,C\}$, and $\epsilon \in \{1,-1\}$, $F_{T^tZ^\epsilon}[-1] \neq 0$.*

Proof. By definition 1.2.4 and by 2.6, if $B = b_{k+1}$, then $F_{B^tB} = F_{k+1}$ and by 2.5, $F_{B^tB^{-1}} = Q_{k+1}$. Next note that $F_1[\lambda] = \lambda - 1$, $F_2[\lambda] = \lambda^2 - 3\lambda + 1$ and for $m \geq 3$, $F_m[\lambda] = (\lambda-2)F_{m-1}[\lambda] - F_{m-2}[\lambda]$ (see 2.6). Since char $(\mathbb{F}) = 3$, $F_m[-1] = -F_{m-2}[-1]$. Hence

(i) $$F_m[-1] \neq 0 \quad \text{for all } m \geq 1.$$

Further, for $m \geq 1$, $Q_m[-1] = (-1)^m\{1-(-1)+(-1)^2-(-1)^3+\cdots\} = (-1)^m(m+1)$. Hence

(ii) $$Q_m[-1] = (-1)^m(m+1), \quad \text{for all } m \geq 1.$$

Now (i) and (ii) imply (1).

For (2), let $\beta_1,\beta_2,\dots,\beta_k,\gamma_1,\gamma_2,\dots,\gamma_k \in \mathbb{F}^*$. Let $B = b_{k+1}(\beta_1,\beta_2,\dots,\beta_k)$, $b = b_k(\beta_2,\beta_3,\dots,\beta_k)$, $C = b_{k+1}(\gamma_1,\gamma_2,\dots,\gamma_k)$, $c = b_k(\gamma_2,\gamma_3,\dots,\gamma_k)$ and for $1 \leq i \leq 4$, $b_i = b_{k-1}(\beta_{i+2},\dots,\beta_k)$ and $c_i = b_{k-i}(\gamma_{i+2},\dots,\gamma_k)$. We claim that

(iii) $$F_{(C^tB)^{-1}}[-1] = (1-\beta_1\gamma_1)F_{(c^tb)^{-1}}[-1] - \beta_1\gamma_1 F_{(c_1^tb_1)^{-1}}[-1].$$

(iv) $$F_{C^tB^{-1}}[-1] = F_{c^tb^{-1}}[-1] - \beta_1\gamma_1 F_{c_1^tb_1^{-1}}[-1].$$

(v) $$\text{If } \beta_1\gamma_1 = 1, \text{ then } F_{(C^tB)^{-1}}[-1] = -F_{(c_1^tb_1)^{-1}}[-1].$$

(vi) $$\text{If } \beta_1\gamma_1 = 1, \text{ then } F_{C^tB^{-1}}[-1] = -\beta_2\gamma_2 F_{c_2^tb_2^{-1}}[-1].$$

Indeed, (iii) follows from 2.12.2, and (iv) follows from 2.12.3. (v) follows from (iii). For (vi), note that by (iv), $F_{c^t b^{-1}}[-1] = F_{c_1^t b_1^{-1}}[-1] - \beta_2\gamma_2 F_{c_2^t b_2^{-1}}[-1]$. Thus, by (iv) again,

$$\begin{aligned} F_{C^t B^{-1}}[-1] &= F_{c^t b^{-1}}[-1] - F_{c_1^t b_1^{-1}}[-1] \\ &= F_{c_1^t b_1^{-1}}[-1] - \beta_2\gamma_2 F_{c_2^t b_2^{-1}}[-1] - F_{c_1^t b_1^{-1}}[-1] \\ &= -\beta_2\gamma_2 F_{c_2^t b_2^{-1}}[-1]. \end{aligned}$$

Let now B and C be as in (2). Then $\beta_1\gamma_1 = \beta_3\gamma_3 = 1$, so applying (v) twice, we see that $F_{(C^t B)^{-1}}[-1] = -F_{(c_1^t b_1)^{-1}}[-1] = F_{(c_3^t b_3)^{-1}}[-1] = \overline{F}_{k-3}[-1]$, where the last equality follows from the fact that $c_3 = b_3 = b_{k-3}$. Note now that (by 2.7.1), $\overline{F}_{k-3}[-1] = F_{k-3}[-1]$, so by (i), $\overline{F}_{k-3}[-1] \neq 0$, and hence $F_{(C^t B)^{-1}}[-1] \neq 0$. Next, by (vi), $F_{C^t B^{-1}}[-1] = -F_{c_2^t b_2^{-1}}[-1] = -\{F_{c_3^t b_3^{-1}}[-1] - \beta_4\gamma_4 F_{c_4^t b_4^{-1}}[-1]\} = -\{Q_{k-3}[-1] + Q_{k-4}[-1]\} = -\{(-1)^{k-3}(k-2) + (-1)^{k-4}(k-3)\} \in \{1, -1\}$. (Note that this also works when $k = 4$ and 5, where $-F_{c_2^t b_2^{-1}}[-1]$ can be easily computed.) This completes the proof of (2).

5.7. (1) *There are at least* $q - 2 - [\frac{q-2}{2}]$ *elements* $\delta \in \mathbb{K}$ *such that the polynomial* $x^2 - \delta x + \delta$ *is irreducible over* $\mathbb{K}$.

(2) *If* $\delta \in \mathbb{K}$ *is as in* (1) *and* $\alpha \in \mathbb{F}$ *is a root of the polynomial* $x^2 - \delta x + \delta$, *then* $\delta = \alpha^{q+1} = \alpha + \alpha^q$.

Proof. Consider the set of polynomials $P := \{x^2 - \delta x + \delta : \delta \in \mathbb{K}\}$. There are q polynomials in P. For $\delta \in \mathbb{K}$, denote $p_\delta = x^2 - \delta x + \delta$. For $p \in P$, let $r(p)$ be the set of roots of p. Note that for $0, 4 \neq \delta \in \mathbb{K}$, $|r(p_\delta)| = 2$ and if $\gamma, \delta \in \mathbb{K}$ are distinct, then $r(p_\gamma) \cap r(p_\delta) = \emptyset$. Hence if t is the number of polynomials $p_\delta \in P$ such $\delta \neq 0, 4$ and p_δ has a root in $\mathbb{K}$, then $2t + 2 \leq q$, so $t \leq [\frac{q-2}{2}]$. Thus $|\{\delta \in \mathbb{K} : p_\delta$ has a root in $\mathbb{K}\}| \leq [\frac{q-2}{2}] + 2$, and (1) follows.

Let $\delta \in \mathbb{K}$ as in (1). Let α be a root of p_δ in $\mathbb{F}$. Then the other root of p_δ is α^q so $p_\delta = (x - \alpha)(x - \alpha^q)$ and hence $\delta = \alpha^{q+1} = \alpha^q + \alpha$.

Notation 5.8. (1) We denote $\Xi = \{\alpha \in \mathbb{F} - \mathbb{K} : \alpha + \alpha^q = \alpha^{q+1}\}$.

(2) We denote by $\mathcal{D} = \{\delta \in \mathbb{K} : p_\delta[\lambda] = \lambda^2 - \delta\lambda + \delta$ is irreducible over $\mathbb{K}\}$.

5.9. *Set* $u = u(\alpha, \theta)$ *and* $w = u(\beta, \rho)$. *Then:*

$$(1) \qquad s(u) = \begin{bmatrix} 1 & 0 & 0 \\ \alpha & 1 & 0 \\ \theta & \alpha^q & 1 \end{bmatrix} \qquad (s(u))^t = \begin{bmatrix} 1 & \alpha & \theta \\ 0 & 1 & \alpha^q \\ 0 & 0 & 1 \end{bmatrix},$$

$$(2) \qquad (s(u))^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -\alpha & 1 & 0 \\ \theta^q & -\alpha^q & 1 \end{bmatrix} \qquad ((s(u))^{-1}) = \begin{bmatrix} 1 & -\alpha & \theta^q \\ 0 & 1 & -\alpha^q \\ 0 & 0 & 1 \end{bmatrix},$$

$$s(uw^t) = \begin{bmatrix} 1 & \beta & \rho \\ \alpha & \alpha\beta + 1 & \alpha\rho + \beta^q \\ \theta & \beta\theta + \alpha^q & \theta\rho + \alpha^q\beta^q + 1 \end{bmatrix}, \tag{3}$$

$$s((uw^t)^{-1}) = \begin{bmatrix} 1 + \alpha\beta + \theta^q\rho^q & \bar{\beta} - \alpha^q\rho^q & \rho^q \\ -\alpha - \beta^q\theta^q & \alpha^q\beta^q + 1 & -\beta^q \\ \theta^q & -\alpha^q & 1 \end{bmatrix}, \tag{4}$$

$$s(u^{-1}w^t) = \begin{bmatrix} 1 & \beta & \rho \\ -\alpha & 1 - \alpha\beta & -\alpha\rho + \beta^q \\ \theta^q & \beta\theta^q - \alpha^q & \rho\theta^q - \alpha^q\beta^q + 1 \end{bmatrix}, \tag{5}$$

$$s((u^{-1}w^t)^{-1}) = \begin{bmatrix} 1 - \alpha\beta + \theta\rho^q & -\beta + \alpha^q\rho^q & \rho^q \\ \alpha - \theta\beta^q & 1 - \alpha^q\beta^q & -\beta^q \\ \theta & \alpha^q & 1 \end{bmatrix}. \tag{6}$$

Proof. (1) is obvious. For (2), observe that $u^{-1} = u_0(-\theta)u_{k+1}(-\alpha^q)u_k(-\alpha)$, so

$$\begin{aligned} s(u^{-1}) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\theta & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ -\alpha & 1 & 0 \\ \alpha^{q+1} & -\alpha^q & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ -\alpha & 1 & 0 \\ \theta^q & -\alpha^q & 1 \end{bmatrix}. \end{aligned}$$

For (3) and (4), we compute:

$$\begin{aligned} s(uw^t) &= \begin{bmatrix} 1 & 0 & 0 \\ \alpha & 1 & 0 \\ \theta & \alpha^q & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & \beta & \rho \\ 0 & 1 & \beta^q \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & \beta & \rho \\ \alpha & \alpha\beta + 1 & \alpha\rho + \beta^q \\ \theta & \beta\theta + \alpha^q & \theta\rho + \alpha^q\beta^q + 1 \end{bmatrix}. \\ s((uw^t)^{-1} &= \begin{bmatrix} 1 & -\beta & \rho^q \\ 0 & 1 & -\beta^q \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ -\alpha & 1 & 0 \\ \theta^q & -\alpha^q & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 + \alpha\beta + \theta^q\rho^q & \bar{\beta} - \alpha^q\rho^q & \rho^q \\ -\alpha - \beta^q\theta^q & \alpha^q\beta^q + 1 & -\beta^q \\ \theta^q & -\alpha^q & 1 \end{bmatrix}. \end{aligned}$$

For (5) and (6) we compute:

$$s(u^{-1}w^t) = \begin{bmatrix} 1 & 0 & 0 \\ -\alpha & 1 & 0 \\ \theta^q & -\alpha^q & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & \beta & \rho \\ 0 & 1 & \beta^q \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \beta & \rho \\ -\alpha & 1-\alpha\beta & -\alpha\rho+\beta^q \\ \theta^q & \beta\theta^q-\alpha^q & \rho\theta^q-\alpha^q\beta^q+1 \end{bmatrix}$$

$$s((u^{-1}w^t)^{-1}) = \begin{bmatrix} 1 & -\beta & \rho^q \\ 0 & 1 & -\beta^q \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ \alpha & 1 & 0 \\ \theta & \alpha^q & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1-\alpha\beta+\theta\rho^q & -\beta+\alpha^q\rho^q & \rho^q \\ \alpha-\theta\beta^q & 1-\alpha^q\beta^q & -\beta^q \\ \theta & \alpha^q & 1 \end{bmatrix}.$$

5.10. *Let $X = X(\alpha, \theta)$ and $Y = X(\beta, \rho)$. Then:*

(1) $\alpha\,(F_{XY^t}, 1) = \binom{k+1}{2} + (\alpha\beta + \theta^q\rho^q + 2)k + \alpha^q\beta^q$.
(2) $\alpha\,(F_{X^{-1}Y^t}, 1) = 3 - \alpha\beta - \alpha^q\beta^q + \theta\rho^q$.
(3) *If* $\alpha\,(F_{XY^t}, 1) = 0$, *then* $(\alpha^q\beta^q - \alpha\beta)(k-1) = (\theta^q\rho^q - \theta\rho)k$.
(4) *If* $\alpha\,(F_{XX^t}, 1) = 0$, *then* $(\alpha^{2q} - \alpha^2)(k-1) = (\theta^{2q} - \theta^2)k$. *Further, if* $\alpha \in \Xi$, *then* $(\alpha^q - \alpha)(k-1) = (\theta^q - \theta)k$.

Suppose further that $\alpha \in \Xi$, *and set* $\delta = \alpha + \alpha^q = \alpha^{q+1}$. *Then:*

(5) *If* $\alpha(F_{X^{-1}X^t}, 1) = 0$, *then* $\delta^2 - 2\delta = 3 + \theta^{q+1}$.
(6) *If* $\theta = \alpha$, *then* $\alpha(F_{XX^t}, 1) \neq 0$, *while if* $\theta = \alpha^q$, *then either* $\alpha(F_{XX^t}, 1) \neq 0$, *or* $2k - 1 \equiv 0 \pmod{\mathrm{char}(\mathbb{K})}$ *and* $8\delta^2 - 16\delta + 11 = 0$.
(7) *If* $\theta = \alpha$ *or* α^2, *then either* $\alpha(F_{X^{-1}X^t}, 1) \neq 0$, *or* $\delta^2 - 3\delta - 3 = 0$.
(8) *Suppose* char $(\mathbb{K}) \neq 2$. *Suppose further that* $\beta = \alpha^q$, $\rho = \theta$ *and* $\theta \neq \mathbb{K}$, *then for* $\{T, Z\} = \{X, Y\}$, $\alpha(F_{TZ^t}, 1) \neq 0$.
(9) *If* $\beta = \alpha^q$, $\rho = \theta$ *and* $2\delta \neq 3 + \theta^{q+1}$, *then for* $\{T, Z\} = \{X, Y\}$, $\alpha(F_{T^{-1}Z^t}, 1) \neq 0$.
(10) *We can choose* $\alpha \in \Xi$ *and* $\theta \in \mathbb{F} - \mathbb{K}$, *with* $\theta + \theta^q = \alpha^{q+1} = \delta$, *such that if we set* $X = X(\alpha, \theta)$ *and* $Y = X(\alpha^q, \theta)$, *then either*
(10i) $q = 2, \theta = \alpha$, *and for* $\epsilon \in \{1, -1\}$, *and* $Z \in \{X, Y\}, \alpha(F_{Z^\epsilon Z^t}, 1) \neq 0$. *Or*
(10ii) $q = 4, \theta = \alpha + 1$ *and there exists* $\beta \in \mathbb{F} - \{\alpha, \alpha^q\}$, *with* $\beta^{q+1} = \delta$, *such that if we set* $W = X(\beta, \theta)$, *then for* $\epsilon \in \{1, -1\}$, *and* $Z \in \{X, Y, W\}, \alpha(F_{Z^\epsilon Z^t}, 1) \neq 0$. *Or*
(10iii) $q \neq 2, 4$ *and* $\alpha(F_{T^\epsilon Z^t}, 1) \neq 0$, *for* $T, Z \in \{X, Y\}$ *and* $\epsilon \in \{1, -1\}$.

Proof. Set $u = u(\alpha, \theta)$ and $w = u(\beta, \rho)$. For (1), let $x_k = a_k^t a_k$, $y_k = b_k b_k^t$, and $g = \mathrm{diag}\,(x_k, 1, y_k^{-1})$. Note that $F_{XY^t} = F_{Y^tX}$. Further $Y^tX = w^t\mathrm{diag}\,(a_k^t, 1, (b_k^{-1})^t)\mathrm{diag}\,(a_k, 1, b_k^{-1})u$. Thus, clearly, $F_{XY^t} = F_{hg}$, where

$h = uw^t$. By 5.9.4, $h^{-1} = \operatorname{diag}(I_{k-1}, s, I_{k-1})$, with

$$s = s((uw^t)^{-1}) = \begin{bmatrix} 1+\alpha\beta+\theta^q\rho^q & \bar{\beta}-\alpha^q\rho^q & \rho^q \\ -\alpha-\beta^q\theta^q & \alpha^q\beta^q+1 & -\beta^q \\ \theta^q & -\alpha^q & 1 \end{bmatrix}.$$

Thus, by 2.10 (with $A = x_k$ and $B = y_k^{-1}$), $\alpha(F_{hg}, 1) = \alpha(R[\lambda], 1)$, where

(i) $$R[\lambda] = (\beta_{22}\lambda - 1)F_A F_B - (\beta_{33} - 1)\lambda F_A F_{M_{1,1}(B)} - (\beta_{11} - 1)\lambda F_{M_{k,k}(A)} F_B,$$

and the β_{ij} are given by matrix s above. Using 2.6, we see that

$$F_A = F_k, \quad F_{M_{k,k}(A)} = G_{k-1} \quad \text{and} \quad F_B = \overline{F}_k.$$

Hence (i) implies

$$R[\lambda] = \{(\alpha^q\beta^q + 1)\lambda - 1\}F_k\overline{F}_k - (\alpha\beta + \theta^q\rho^q)\lambda G_{k-1} \cdot \overline{F}_k.$$

Now 2.6 gives

$$F_k \equiv (-1)^k \left\{1 - \binom{k+1}{2}\lambda\right\} \pmod{(\lambda^2)}$$
$$G_{k-1} \equiv (-1)^{k-1} \left\{k - \binom{k+1}{3}\lambda\right\} \pmod{(\lambda^2)}$$
$$\overline{F}_k \equiv (-1)^k\{1 - (2k-1)\lambda\} \pmod{(\lambda^2)}$$

Hence modulo the ideal (λ^2),

$$\begin{aligned} R[\lambda] &\equiv \{(\alpha^q\beta^q + 1)\lambda - 1\} \cdot \left\{1 - \binom{k+1}{2}\lambda\right\} \cdot \{1 - (2k-1)\lambda\} \\ &\quad + (\alpha\beta + \theta^q\rho^q)\lambda k \\ &\equiv -1 + \left\{\binom{k+1}{2} + (\alpha\beta + \theta^q\rho^q + 2)k + \alpha^q\beta^q\right\}\lambda. \end{aligned}$$

This shows (1).

For (2), let $x_k = a_k^{-1}a_k^t$, $y_k = b_k(b_k^{-1})^t$ and $g = \operatorname{diag}(x_k, 1, y_k)$. Using 5.2.3, we see that $X^{-1}Y^t = u^{-1}\operatorname{diag}(a_k^{-1}, 1, b_k)w^t\operatorname{diag}(a_k^t, 1, (b_k^{-1})^t) = u^{-1}w^t\operatorname{diag}(a_k^{-1}, 1, b_k)\operatorname{diag}(a_k^t, 1, (b_k^{-1})^t) = hg$, where $h = u^{-1}w^t$. Thus, $F_{X^{-1}Y^t} = F_{hg}$. By 5.9.6, $h^{-1} = \operatorname{diag}(I_{k-1}, s, I_{k-1})$, with

$$s = s((u^{-1}w^t)^{-1}) = \begin{bmatrix} 1-\alpha\beta+\theta\rho^q & -\beta+\alpha^q\rho^q & \rho^q \\ \alpha-\theta\beta^q & 1-\alpha^q\beta^q & -\beta^q \\ \theta & \alpha^q & 1 \end{bmatrix}.$$

Using 2.10 again (with $A = x_k$ and $B = y_k$), $\alpha(F_{hg}, 1) = \alpha(R[\lambda], 1)$, with $R[\lambda]$ as in (i) and the β_{ij} are given by the matrix s above. Using 2.5 and 2.11, we see that

$$F_A = Q_k, \quad F_{M_{k,k}(A)} = Q_{k-1}, \quad F_B = Q_k.$$

Hence

$$R[\lambda] = \{(1 - \alpha^q\beta^q)\lambda - 1\}Q_k^2 - (-\alpha\beta + \theta\rho^q)\lambda Q_{k-1} \cdot Q_k.$$

Now

$$Q_k \equiv (-1)^k(1-\lambda) \pmod{(\lambda^2)}$$
$$Q_{k-1} \equiv (-1)^{k-1} \pmod{(\lambda)}.$$

Hence modulo the ideal (λ^2),

$$\begin{aligned} R[\lambda] &\equiv \{(1 - \alpha^q\beta^q)\lambda - 1\}(1-\lambda)^2 + (-\alpha\beta + \theta\rho^q)\lambda \\ &\equiv -1 + \{3 - \alpha\beta - \alpha^q\beta^q + \theta\rho^q\}\lambda. \end{aligned}$$

This shows (2).

Suppose $\alpha(F_{XY^t}, 1) = 0$. Applying σ_q, we get

$$\alpha(F_{XY^t}, 1) = 0 = \alpha(F_{XY^t}, 1)\sigma_q,$$

hence

$$(\alpha\beta + \theta^q\rho^q)k + \alpha^q\beta^q = (\alpha^q\beta^q + \theta\rho)k + \alpha\beta$$

so

$$(\alpha^q\beta^q - \alpha\beta)(k-1) = (\theta^q\rho^q - \theta\rho)k$$

and (3) is proved. For (4), take $Y = X$ in (3), to get $(\alpha^{2q} - \alpha^2)(k-1) = (\theta^{2q} - \theta^2)k$. Further, $(\alpha^{2q} - \alpha^2) = (\alpha^q + \alpha)(\alpha^q - \alpha)$, and $(\theta^{2q} - \theta^2) = (\theta^q + \theta)(\theta^q - \theta)$. So if $\alpha \in \Xi$, $(\alpha^q + \alpha) = \alpha^{q+1} = \theta^q + \theta$. This shows (4).

From now on assume $\alpha \in \Xi$ and set $\delta = \alpha^{q+1}$. For (5), take $X = Y$ in (2) and note that $\alpha^2 + \alpha^{2q} = (\alpha + \alpha^q)^2 - 2\alpha^{q+1} = \delta^2 - 2\delta$.

Suppose $\theta = \alpha$ and $\alpha(F_{XX^t}, 1) = 0$. Then, by (4), $(\alpha^q - \alpha)(k-1) = (\alpha^q - \alpha)k$. Hence $\alpha^q = \alpha$, which is false, since $\alpha \notin \mathbb{K}$. Suppose $\theta = \alpha^q$ and $\alpha(F_{XX^t}, 1) = 0$. Then, by (4), $(\alpha^q - \alpha)(k-1) = (\alpha - \alpha^q)k$ hence $(2k-1)(\alpha^q - \alpha) = 0$. As above, we get $2k - 1 = 0$ in $\mathbb{K}$, so $\binom{k+1}{2} = \frac{3}{8}$ in $\mathbb{K}$. Also, by (1), $0 = \alpha(F_{XX^t}, 1) = \binom{k+1}{2} + (\alpha^2 + \theta^{2q} + 2)k + \alpha^{2q} = \frac{3}{8} + (\alpha^2 + \alpha^2 + 2)\frac{1}{2} + \alpha^{2q} = \frac{11}{8} + \alpha^2 + \alpha^{2q}$. Since $\alpha^2 + \alpha^{2q} = \delta^2 - 2\delta$, we get that $\frac{11}{8} + \delta^2 - 2\delta = 0$. This shows (6).

For (7) suppose that $\theta = \alpha$ or α^q. Then, $\theta^{q+1} = \delta$, so, by (5), if $\alpha(F_{X^{-1}X^t}, 1) = 0$, then $\delta^2 - 2\delta = 3 + \delta$, and $\delta^2 - 3\delta - 3 = 0$, this shows (7).

Assume the hypothesis of (8). Note that $\alpha^q\beta^q - \alpha\beta = 0$. Thus, by (3), if $\alpha(F_{XY^t}, 1) = 0$, then $0 = (\theta^{2q} - \theta^2)k = \delta(\theta^q - \theta)k$. Thus since $\delta \neq 0$ and since we are assuming that $\theta \notin \mathbb{K}$, $k = 0$, in $\mathbb{K}$. Then, (1) implies that $\delta = 0$, a contradiction. By symmetry, (8) holds.

Assume the hypothesis of (9). Note again that $\alpha\beta = \delta$, so (9) follows immediately from (2).

For (10), assume $Y = X(\alpha^q, \theta)$. Suppose first that $\mathrm{char}(\mathbb{K}) = 2$. Note that by (4):

(ii) If α, $\theta \in \mathbb{F} - \mathbb{K}$ such that $\theta + \theta^q = \alpha^{q+1}$, then for $X = X(\alpha, \theta)$, $\alpha(F_{XX^t}, 1) \neq 0$.

This is because (4) implies that if k is odd and $\alpha(F_{XX^t}, 1) = 0$, then $\theta^q + \theta = 0$, while if k is even and $\alpha(F_{XX^t}, 1) = 0$, then $\alpha^q + \alpha = 0$.

For $q = 2$, take $\delta = 1$, for $q > 2$, pick $1 \neq \delta \in \mathcal{D}$ (note that this is possible by 5.7). Further, if $q > 4$, take δ such that $\delta^2 + \delta + 1 \neq 0$ (note that this is possible). Let $\alpha \in \Xi$, with $\alpha^{q+1} = \delta$. If $q = 2$, take $\theta = \alpha$, if $q = 4$, take $\theta = \alpha + 1$ and if $q > 4$, take $\theta = \alpha + \delta$. Note that $\theta \notin \mathbb{K}$. When $q = 4$, we take $W = X(\beta, \theta)$, with $\beta \in \mathbb{F} - (\mathbb{K} \cup \{\alpha, \alpha^q\})$, such that $\beta^{q+1} = \alpha^{q+1} = \delta$. Note that such a choice of β is possible. Now, by (ii), for all $q \geq 2$, $\alpha(F_{ZZ^t}, 1) \neq 0$, for $Z \in \{X, Y, W\}$.

Next, for $q = 4$, $\theta^{q+1} = (\alpha+1)^{q+1} = (\alpha^q+1)(\alpha+1) = \alpha^{q+1} + (\alpha^q+\alpha) + 1 = 1$. Of course, when $q = 2$, $\theta^{q+1} = 1$. Also, by (2), for $Z \in \{X, Y, W\}$, if $Z = X(\gamma, \theta)$, then $\alpha(F_{Z^{-1}Z^t}, 1) = 3 + \gamma^2 + \gamma^{2q} + 1 = \gamma^2 + \gamma^{2q} = (\gamma + \gamma^q)^2$. Since $\gamma \notin \mathbb{K}$, for all possibilities of γ and for $q = 2, 4$, $\alpha(F_{Z^{-1}Z^t}, 1) \neq 0$. Thus (10i) and (10ii) are proved.

We now assume that $\mathrm{char}(\mathbb{F}) = 2$ and $q > 4$. Now $\theta^{q+1} = (\alpha + \delta)^{q+1} = (\alpha^q+\delta)(\alpha+\delta) = \alpha^{q+1} + \delta(\alpha^q+\alpha) + \delta^2 = \delta + \delta^2 + \delta^2 = \delta$. Hence, $3 + \theta^{q+1} = \delta + 1$. So if $\delta^2 - 2\delta = 3 + \theta^{q+1}$, then $\delta^2 = \delta + 1$, this contradicts the choice of δ (recall $\delta^2 + \delta + 1 \neq 0$). Hence, by (5), $\alpha(F_{Z^{-1}Z^t}, 1) \neq 0$, for $Z \in \{X, Y\}$.

Suppose $\alpha(F_{XY^t}, 1) = 0$. Then, by (3), (with $\beta = \alpha^q$), we get $0 = (\theta^q + \theta)^2 k = \delta^2 k$, so $k \equiv 0 \pmod 2$. Then by (1), $\binom{k+1}{2} + \delta = 0$. Thus $k \equiv 2 \pmod 4$ (since $\delta \neq 0$) and $\delta = 1$, contradicting the choice of δ. Thus $\alpha(F_{XY^t}, 1) \neq 0$; by symmetry, $\alpha(F_{YX^t}, 1) \neq 0$.

Next note that we showed that $\theta^{q+1} = \delta$. Thus $\theta^{q+1} + 3 = \delta + 1$. Since $\delta \neq 1$, $\delta + 1 \neq 0$, so by (9), $\alpha(F_{T^{-1}Z^t}, 1) = 0$, for $\{T, Z\} = \{X, Y\}$. Thus (10iii) holds in case $\mathrm{char}(\mathbb{K}) = 2$.

So suppose $\mathrm{char}(\mathbb{K}) \neq 2$. Suppose further that $q \neq 5$. We take $3 \neq \delta \in \mathcal{D}$, $\alpha \in \Xi$, with $\alpha^{q+1} = \delta$ and $\theta = \alpha$. Since $\theta \notin \mathbb{K}$, (8) implies that for $\{T, Z\} = \{X, Y\}$, $\alpha(F_{TZ^t}, 1) \neq 0$. Since $\delta \neq 3$, (and $\theta^{q+1} = \alpha^{q+1} = \delta$), (9) implies that for $\{T, Z\} = \{X, Y\}$, $\alpha(F_{T^{-1}Z^t}, 1) \neq 0$. Next we show that we can pick $\delta \in \mathcal{D}$, such that

(iii) $$\delta \neq 3 \quad \text{and} \quad 8\delta^2 - 16\delta + 11 \neq 0 \neq \delta^2 - 3\delta - 3.$$

By (6) and (7), this shows (10), for $q \neq 5$. If $q \geq 13$, then, by 5.7.1, $|\mathcal{D}| \geq 6$, so clearly, we can pick $\delta \neq 3$ such that (iii) holds. So suppose $q \leq 11$. Suppose $\mathrm{char}(\mathbb{K}) = 3$. Then $\delta^2 - 3\delta - 3 \neq 0$, so if $q = 9$, then, by 5.7.1, we can pick δ ($\neq 3$) so that $8\delta^2 - 16\delta + 11 \neq 0$, while if $q = 3$, take $\delta = -1$, so (iii) holds in this case. For $q = 11$, take $\delta = 1$. For $q = 7$, take $\delta = 2$.

Finally, suppose $q = 5$. We take $\delta = 1$, $\alpha \in \Xi$, with $\alpha^{q+1} = \delta$ and we let θ be as follows. If $k \not\equiv 2 \pmod 5$, $\theta = \theta_1 = \alpha + 3(\alpha - \alpha^q) = 2\alpha^q - \alpha$, while if $k \equiv 2 \pmod 5$, $\theta = \theta_1^q$ (note that $\theta + \theta^q = \alpha + \alpha^q = \alpha^{q+1}$). Note that if $\theta \in \mathbb{K}$, then $\alpha \in \mathbb{K}$, which is false. Thus $\theta \notin \mathbb{K}$. Hence, by (8), for $\{T, Z\} = \{X, Y\}$, $\alpha(F_{TZ^t}, 1) \neq 0$. Next, $\theta^{q+1} = (2\alpha - \alpha^q)(2\alpha^q - \alpha) = 4\delta - 2(\alpha^2 + \alpha^{2q}) + \delta = -2(\delta^2 - 2\delta)$. Thus

$$\text{(iv)} \qquad \theta^{q+1} = 2.$$

By (iv) $\theta^{q+1} + 3 = 0 \neq 2\delta$. Hence, by (9), for $\{T, Z\} = \{X, Y\}$, $\alpha(F_{T^{-1}Z^t}, 1) \neq 0$. Also, $\delta^2 - 2\delta = -1 \neq 0 = 3 + \theta^{q+1}$, so, by (5), $\alpha(F_{Z^{-1}Z^t}, 1) \neq 0$, for $Z \in \{X, Y\}$.

Next, $\theta_1^q - \theta_1 = 2(\alpha - \alpha^q) - (\alpha^q - \alpha) = 3(\alpha - \alpha^q) = 2(\alpha^q - \alpha)$. So:

$$\text{(v)} \qquad \begin{aligned} &\text{If } k \not\equiv 2 \pmod 5,\ \theta^q - \theta = 2(\alpha^q - \alpha) \\ &\text{and if } k \equiv 2 \pmod 5,\ \theta^q - \theta = 3(\alpha^q - \alpha). \end{aligned}$$

Suppose first that $k \not\equiv 2 \pmod 5$. Suppose $\alpha(F_{XX^t}, 1) = 0$, then by (4) and (v), $(k - 1) = 2k$ so $k \equiv -1 \pmod 5$. Then, by (1), $\alpha(F_{XX^t}, 1) = \binom{k+1}{2} + (\alpha^2 + \theta^{2q} + 2)k + \alpha^{2q} = -(\alpha^2 + \theta^{2q} + 2) + \alpha^{2q} = \alpha^{2q} - \alpha^2 - \theta^{2q} - 2 = \alpha^{2q} - \alpha^2 - (2\alpha - \alpha^q)^2 - 2 = \alpha^{2q} - \alpha^2 - 4\alpha^2 + 4 - \alpha^{2q} - 2 = 2 \neq 0$, a contradiction.

Suppose $\alpha(F_{YY^t}, 1) = 0$. Then, by (4), and (v) (replacing α by α^q in (4)), $-(k - 1) \equiv 2k \pmod 5$, so $k \equiv 2 \pmod 5$, a contradiction.

Finally, suppose $k \equiv 2 \pmod 5$. Then, by (1), $\alpha(F_{XX^t}, 1) = \binom{k+1}{2} + (\alpha^{2q} + \theta^{2q} + 2)k + \alpha^2 = 3 + 2(\alpha^{2q} + \theta^{2q} + 2) + \alpha^2 = 2 + 2\alpha^{2q} + \alpha^2 + 2\theta^{2q} = 2 + 2\alpha^{2q} + \alpha^2 + 2(2\alpha^q - \alpha)^2 = 2 + 2\alpha^{2q} + \alpha^2 + 2(4\alpha^{2q} - 4 + \alpha^2) = -1 + 3\alpha^2 \neq 0$.

Suppose $\alpha(F_{YY^t}, 1) = 0$. Then, by (4), and (v) (replacing α by α^q in (4)), $-(k - 1) \equiv 3k \pmod 5$, so $k \equiv -1 \pmod 5$, a contradiction. This completes the proof of (10) and of 5.10.

5.11. *Let $\beta_1, \dots, \beta_{k-1}, \gamma_1, \dots, \gamma_{k-1} \in \mathbb{K}^*$. Let also $\alpha, \theta, \beta, \rho \in \mathbb{F}^*$ such that $\alpha^{q+1} = \theta + \theta^q$, $\beta^{q+1} = \rho + \rho^q$. Set $a = a_k(\beta_1, \dots, \beta_{k-1})$, $a_1 = a_k(\gamma_1, \dots, \gamma_{k-1})$, $b = b_k(\beta_1, \dots, \beta_{k-1})$, $b_1 = b_k(\gamma_1, \dots, \gamma_{k-1})$, $g = \operatorname{diag}(a, 1, b^{-1})$, $g_1 = \operatorname{diag}(a_1, 1, b_1^{-1})$, $B = u_1^{k+1}(-\alpha^q)\operatorname{diag}(1, b)$, $B_1 = u_1^{k+1}(-\beta^q)\operatorname{diag}(1, b_1)$, $u = u(\alpha, \theta)$, $w = u(\beta, \rho)$, $X = gu$ and $Y = g_1 w$. Finally let $\epsilon \in \{-1, 1\}$. Then:*

(1) *$XY^t = guw^t g_1^t$, $(XY^t)^{-1} = (g_1^t)^{-1}(uw^t)^{-1}g^{-1}$.*

(2) *$X^{-1}Y^t = u^{-1}w^t g^{-1}g_1^t$ and $(X^{-1}Y^t)^{-1} = (g_1^t)^{-1}g(u^{-1}w^t)^{-1}$.*

(3) *$X = \begin{bmatrix} a & 0_{k,k+1} \\ E & B^{-1} \end{bmatrix}$ with E some $(k + 1) \times k$ matrix, such that $E_{1,k} = \alpha \neq 0$.*

(4)

$$X^\epsilon Y^t = \begin{bmatrix} a^\epsilon a_1^t & R \\ S & T \end{bmatrix} \qquad (X^\epsilon Y^t)^{-1} = \begin{bmatrix} T' & R' \\ S' & B_1^t B^\epsilon \end{bmatrix}$$

with T', T, R, R', S, S' some $k\times k$, $(k+1)\times(k+1)$, $k\times(k+1)$, $k\times(k+1)$, $(k+1)\times k$, $(k+1)\times k$, matrices respectively. Further, the first $k-1$ rows of R are zero.

(5) *Let $S \in \{Y^t, X^\epsilon Y^t\}$. Then for $1 \le i \le k-1$, $v_i S = v + \delta_{i+1} v_{i+1}$, with $v \in \mathcal{V}_i$ and $\delta_{i+1} \in \mathbb{K}^*$.*

(6) *$S_{k,n} \neq 0$, for all $S \in \{Y^t, X^\epsilon Y^t\}$.*

(7) *For $S \in \{Y^t, X^\epsilon Y^t\}$, there exists $v \in \mathcal{V}_k$, $\eta \in \mathbb{F}$ and $\mu \in \mathbb{F}^*$ such that:*

(7i) $$v_{k+1}S^{-1} \equiv \eta v_{k+1} + \mu v_{k+2} \pmod{\mathcal{V}_k}.$$

(7ii) $$vS^{-1} \equiv (\eta + \rho^{1-q})v_{k+1} + \mu v_{k+2} \pmod{\mathcal{V}_k}.$$

(7iii) *In all cases $\mu = -\beta^q$. If $S = Y^t$, $\eta = 1$, while if $S = X^\epsilon Y^t$, $\eta = 1 + \epsilon\alpha^q\beta^q$.*

(8) *For $S \in \{Y^t, X^\epsilon Y^t\}$, $V = \langle \mathcal{O}(v_1, S)\rangle$ iff $-\rho^{1-q}$ is not a root of F_Z, where $Z = B_1^t$, if $S = Y^t$ and $Z = B_1^t B^\epsilon$, if $S = X^\epsilon Y^t$.*

(9) *If $\beta \neq 0$, then $V = \langle \mathcal{O}(v_1, Y^t)\rangle$.*

Proof. (1) is obvious. For (2), we have $X^{-1}Y^t = u^{-1}g^{-1}w^t g_1^t$. By 5.2.3, $[g^{-1}, w^t] = 1$, and (2) follows. For (3) recall from 5.4.1 that

$$u = u_{k+1}(\alpha^q)u_k(\alpha)u_0(-\theta^q).$$

Further by 5.2.2, $gu_{k+1}(\alpha^q) = \mathrm{diag}\,(a, B^{-1})$. Thus

$$X = \mathrm{diag}\,(a, B^{-1})u_k(\alpha)u_0(-\theta^q).$$

Note now that

$$s(u_k(\alpha)u_0(-\theta^q)) = \begin{bmatrix} 1 & 0 & 0 \\ \alpha & 1 & 0 \\ -\theta^q & 0 & 1 \end{bmatrix}.$$

Hence (3) follows, the $(1,k)$-entry of E is $\alpha(B^{-1})_{1,1} - \theta^q(B^{-1})_{1,2} = \alpha\cdot 1 - \theta^q\cdot 0 = \alpha$.

To show (4) and (5), let $1 \le i \le k-1$, then $v_i u^{-1}w^t = v_i$, so $v_i X^{-1}Y^t = v_i g^{-1}g_1^t$. Also $v_i g \in \mathcal{V}_i$, so $v_i g(uw^t) = v_i g$ and $v_i XY^t = v_i g g_1^t$. We conclude that:

(i) $$\text{For } 1 \le i \le k-1,\ v_i X^\epsilon Y^t = v_i g^\epsilon g_1^t.$$

Now the shape of $X^\epsilon Y^t$ follows from (3) and (i), since, by (i), the first $k-1$ rows of R are zero. Also the shape of $(X^\epsilon Y^t)^{-1}$, follows from (3). For (5), we use (i). Note that a^ϵ is unipotent, lower triangular and a_1^t is upper triangular unipotent with $(a_1^t)_{i,j} = 0$, for $j > i+1$, and $(a_1^t)_{i,i+1} \neq 0$. This easily implies (5), for $S = X^\epsilon Y^t$. For $S = Y^t$, $v_i Y^t = v_i + \gamma_{k-i}v_{i+1}$, for all $1 \le i \le k-1$, thus (5) holds for Y^t as well.

Recall now that

$$s(uw^t) = \begin{bmatrix} 1 & \beta & \rho \\ \alpha & \alpha\beta+1 & \alpha\rho+\beta^q \\ \theta & \beta\theta+\alpha^q & \theta\rho+\alpha^q\beta^q+1 \end{bmatrix}$$

$$s((uw^t)^{-1}) = \begin{bmatrix} 1+\alpha\beta+\theta^q\rho^q & \bar{\beta}-\alpha^q\rho^q & \rho^q \\ -\alpha-\beta^q\theta^q & \alpha^q\beta^q+1 & -\beta^q \\ \theta^q & -\alpha^q & 1 \end{bmatrix}$$

$$s(u^{-1}w^t) = \begin{bmatrix} 1 & \beta & \rho \\ -\alpha & 1-\alpha\beta & -\alpha\rho+\beta^q \\ \theta^q & \beta\theta^q-\alpha^q & \rho\theta^q-\alpha^q\beta^q+1 \end{bmatrix}$$

$$s((u^{-1}w^t)^{-1}) = \begin{bmatrix} 1-\alpha\beta+\theta\rho^q & -\beta+\alpha^q\rho^q & \rho^q \\ \alpha-\theta\beta^q & 1-\alpha^q\beta^q & -\beta^q \\ \theta & \alpha^q & 1 \end{bmatrix}.$$

Note now that $v_k g^{-1} \equiv v_k \equiv v_k g \pmod{\mathcal{V}_{k-1}}$, $v_{k+1}g^{-1} = v_{k+1}$ and $v_{k+2}g^{-1} = v_{k+2}$. Since $u^\epsilon w^t$ fixes $\mathcal{V}_{k-1}$, we see that,

$$v_k(X^\epsilon Y^t) \equiv v_k(u^\epsilon w^t)g_1^t \pmod{\mathcal{V}_{k-1}}.$$

Thus modulo $\mathcal{V}_k$, $v_k(X^\epsilon Y^t) \equiv (\beta v_{k+1} + \rho v_{k+2})g_1^t \equiv \beta v_{k+1} + \rho(v' + \eta v_n)$, with $v' \in \langle v_{k+2}, \dots, v_{n-1}\rangle$, $\eta \in \mathbb{F}^*$. This is because the $(k,1)$ entry of b_1^{-1} is $\eta = \gamma_1\gamma_2\cdots\gamma_{k-1}$, and $g_1^t = \operatorname{diag}(a_1^t, 1, (b_1^{-1})^t)$. This shows (6), for $S = X^\epsilon Y^t$ and it is easy to see that (6) holds for $S = Y^t$ as well.

Next, modulo $\mathcal{V}_k$, we have $-\rho^{-q}\beta^q v_k(XY^t)^{-1} = -\rho^{-q}\beta^q v_k(uw^t)^{-1}g^{-1} \equiv ((\alpha^q\beta^q+\beta^{q+1}\rho^{-q})v_{k+1}-\beta^q v_{k+2})g^{-1} = (\alpha^q\beta^q+\beta^{q+1}\rho^{-q})v_{k+1}-\beta^q v_{k+2}$. Since $\beta^{q+1} = \rho+\rho^q$, we see that $-\rho^{-q}\beta^q v_k(XY^t)^{-1} \equiv (\alpha^q\beta^q+1+\rho^{1-q})v_{k+1} - \beta^q v_{k+2}$. Note that $v_{k+1}(XY^t)^{-1} \equiv (\alpha^q\beta^q+1)v_{k+1}-\beta^q v_{k+2} \pmod{\mathcal{V}_k}$. This shows (7), for $S = XY^t$.

Let $v \in \mathcal{V}_k$, such that $v(g_1^t)^{-1}g = v_k$. Then, modulo $\mathcal{V}_k$,

$$\begin{aligned} -\rho^{-q}\beta^q v(X^{-1}Y^t)^{-1} &= -\rho^{-q}\beta^q v_k(u^{-1}w^t)^{-1} \\ &\equiv ((\beta^{q+1}\rho^{-q} - \alpha^q\beta^q)v_{k+1} - \beta^q v_{k+2}) \\ &= (1-\alpha^q\beta^q+\rho^{1-q})v_{k+1} - \beta^q v_{k+2}. \end{aligned}$$

Note that $v_{k+1}(X^{-1}Y^t)^{-1} \equiv (1-\alpha^q\beta^q)v_{k+1}-\beta^q v_{k+2} \pmod{\mathcal{V}_k}$. This shows (7), for $S = X^{-1}Y^t$.

Next

$$\begin{aligned} -\rho^{-q}\beta^q v_k(Y^t)^{-1} &= -\rho^{-q}\beta^q v_k(g_1^t)^{-1}(w^t)^{-1} = -\rho^{-q}\beta^q v_k(w^t)^{-1} \\ &= -\rho^{-q}\beta^q v_k + \rho^{-q}\beta^{q+1}v_{k+1} - \beta^q v_{k+2} \\ &= -\rho^{-q}\beta^q v_k + (1+\rho^{1-q})v_{k+1} - \beta^q v_{k+2}. \end{aligned}$$

Also $v_{k+1}(Y^t)^{-1} = v_{k+1} - \beta^q v_{k+2}$, thus (7) holds for $S = Y^t$ and (7) is proved.

For (8), set $\mathcal{W} = \langle \mathcal{O}(v_1, S)\rangle$. Set also $Z = B_1^t$, if $S = Y^t$ and $Z = B_1^t B^\epsilon$, if $S = X^\epsilon Y^t$. By (5), $\mathcal{V}_k \subseteq \mathcal{W}$. Let $\eta, \mu \in \mathbb{F}$ be as in (7iii). Since $\mathcal{V}_k \subseteq \mathcal{W}$,

$$\text{(ii)} \qquad \rho^{1-q} v_{k+1} + \eta v_{k+1} + \mu v_{k+2} \in \mathcal{W}.$$

Also, by (3), (4) and (7i), $v_{k+1}\mathrm{diag}\,(I_k, Z) = \eta v_{k+1} + \mu v_{k+2}$. Thus, by (ii), $\rho^{1-q} v_{k+1} + v_{k+1}\mathrm{diag}\,(I_k, Z) \in \mathcal{W}$, now (8) follows from (4), (5) and 1.17 (taking S^{-1} in place of S in 1.17); note that $\langle \mathcal{O}(v_{k+1}, \mathrm{diag}\,(I_k, Z))\rangle = \langle v_{k+1}, \dots, v_n\rangle$.

Finally, for (9), note that if $\beta \neq 0$, then $\rho^{1-q} \neq -1$, since $0 \neq \beta^{q+1} = \rho + \rho^q$. Since 1 is the only root of $F_{B_1^t}$, $-\rho^{1-q}$ is not a root of $F_{B_1^t}$, so (9) follows from (8).

5.12. *Let $\beta_1, \dots, \beta_{k-1}, \gamma_1, \dots, \gamma_{k-1} \in \mathbb{K}^*$. Let also $\alpha, \theta, \beta, \rho \in \mathbb{F}^*$ such that $\alpha^{q+1} = \theta + \theta^q$, $\beta^{q+1} = \rho + \rho^q$. Set $a = a_k(\beta_1, \dots, \beta_{k-1})$, $a_1 = a_k(\gamma_1, \dots, \gamma_{k-1})$, $b = b_k(\beta_1 \dots, \beta_{k-1})$, $b_1 = b_k(\gamma_1, \dots, \gamma_{k-1})$, $g = \mathrm{diag}\,(a, 1, b^{-1})$, $g_1 = \mathrm{diag}\,(a_1, 1, b_1^{-1})$, $u = u(\alpha, \theta)$, $w = u(\beta, \rho)$, $X = gu$ and $Y = g_1 w$. Finally let $\epsilon \in \{-1, 1\}$.*

Let $S \in \{Y^t, X^\epsilon Y^t\}$ and $R \in \Delta^{\leq 2}(X) \cap \Delta^{\leq 1}(S)$. Then v_1 is a characteristic vector of R.

Proof. The proof is almost identical to the proof of 4.3. Note first that, by 5.11.3, X satisfies the hypotheses of 1.13. Let $h \in \Delta^{\leq 1}(X) \cap \Delta^{\leq 1}(R)$. Then, $[h, X] = 1$, so by 1.13, there exists $0 \neq \beta \in \mathbb{K}$, and $1 \leq r \leq n-1$, such that $h - \beta I_n \in \mathcal{T}_n(r)$. We use 1.15. We take in 1.15, $T = h - \beta I_n$. Note that R commutes with h and hence with T.

Suppose first that $r \leq k-1$, we take in 1.15, $j = r = m$ and $\ell = 0$. Notice that by 5.11.5, hypothesis (a) of 1.15 is satisfied, hypothesis (b) and (c) of 1.15 are satisfied, by definition and we observed that hypothesis (e) of 1.15 is satisfied. Finally, since R centralizes T, $\mathcal{V}_r$ is R-invariant. Hence 1.15 completes the proof in this case.

Suppose next that $r \geq k$, we take in 1.15, $j = k-1$, $\ell = 1$ and $m = k$, if $r = k$ and $m = \dim(\mathrm{im}\,(T))$, if $r > k$. Notice that $\mathcal{V}_m$ is R-invariant. Also, by 5.11.6, $S_{k,n} \neq 0$, so clearly $v_k \notin \ker(ST)$ and hypothesis (c) of 1.15 holds. Thus 1.15 completes the proof in this case too.

5.13. *For $i \in \{1, 2, 3, 4\}$, let $\alpha_i \in \mathbb{F}^*$ and set $B_i = u_1^{k+1}(-\alpha_i)\mathrm{diag}\,(1, b_k)$. Let also $\epsilon \in \{1, -1\}$ and $1 \neq \gamma \in \mathbb{F}^*$. Then:*

1. *If $F_{B_1^t B_2^\epsilon}[\gamma] = 0 = F_{B_3^t B_4^\epsilon}[\gamma]$, then $\alpha_1\alpha_2 = \alpha_3\alpha_4$.*
2. *Suppose $\alpha_1^2 \notin \mathbb{K}$ and $\alpha_2 = \alpha_1^q$. Then γ is a root of at most one of the polynomials $F_{B_1^t B_1^\epsilon}, F_{B_2^t B_2^\epsilon}$ and $F_{B_1^t B_2^\epsilon}$.*
3. *Suppose $\alpha_1^2 \notin \mathbb{K}$ and $\alpha_2 = \alpha_1^q$. Then either we can find $j \in \{1, 2\}$, such that $F_{B_j^t B_j}[\gamma] \neq 0 \neq F_{B_j^t B_j^{-1}}[\gamma]$, or for $\{B, C\} = \{B_1, B_2\}$, $F_{B^t C}[\gamma] \neq 0 \neq F_{B^t C^{-1}}[\gamma]$.*

(4) *If* char$(\mathbb{K}) \neq 2$ *and* $q > 3$*, then we can find* $\alpha_1, \alpha_2 \in \mathbb{K}^*$*, such that* $F_{B_1^t B_2}[-1], F_{B_1^t B_2^{-1}}[-1], F_{B_2^t B_1}[-1], F_{B_2^t B_1^{-1}}[-1]$ *are all distinct from* 0.

(5) *Suppose that* $q = 2$*, and that* $\alpha_1 \notin \mathbb{K}$*. Then,* $F_{B_1^t B_1^{-1}}[\gamma] \neq 0$*. In particular, we can pick* $\alpha_1 \in \mathbb{F} - \mathbb{K}$ *such that* $F_{B_1^t B_1}[\gamma] \neq 0 \neq F_{B_1^t B_1^{-1}}[\gamma]$.

Proof. First observe that, for $1 \leq i \leq 4$, $B_i = b_{k+1}(\alpha_i, 1, \dots, 1)$. We mention that for small values of k ($k = 1, 2$ or 3), direct calculations show (1). For the general case in (1), suppose $F_{B_1^t B_2}[\gamma] = 0 = F_{B_3^t B_4}[\gamma]$. Then, by 2.12.4, $(\gamma - 1)F_k[\gamma] - \alpha_1\alpha_2\gamma G_{k-1}[\gamma] = 0 = (\gamma - 1)F_k[\gamma] - \alpha_3\alpha_4\gamma G_{k-1}[\gamma]$. Suppose $\alpha_1\alpha_2 \neq \alpha_3\alpha_4$. Then $G_{k-1}[\gamma] = 0$, and as $\gamma \neq 1$, $F_k[\gamma] = 0$. This contradicts 2.6.6. Using 2.12.5, it is easy to see that if $F_{B_1^t B_2^{-1}}[\gamma] = 0 = F_{B_3^t B_4^{-1}}[\gamma]$, then $\alpha_1\alpha_2 = \alpha_3\alpha_4$. (2) follows immediately from (1), noticing that $\alpha_1^2, \alpha_1^{2q}$ and α_1^{q+1} are distinct. (3) follows from (2) noticing that, by 2.12.4 and 2.12.5, $F_{B_1^t B_2^\epsilon}[\gamma] = F_{B_2^t B_1^\epsilon}[\gamma]$.

For (4), just choose $\alpha_1, \alpha_2 \in \mathbb{K}^*$ such that -1 is not a root of the polynomial $F_{B_1^t B_2} = F_{B_2^t B_1} = (\lambda - 1)F_k - \alpha_1\alpha_2\lambda G_{k-1}$ nor of the polynomial $F_{B_1^t B_2^{-1}} = F_{B_2^t B_1^{-1}} = (\lambda - 1)Q_k + \alpha_1\alpha_2\lambda Q_{k-1}$, using (1).

For (5), note that as $q = 2$, 2.12.5 shows that, $F_{B_1^t B_1^{-1}}[\lambda] = (\lambda + 1)Q_k + \alpha_1^2\lambda Q_{k-1} = \lambda^{k+1} + 1 + \alpha_1^2\lambda Q_{k-1}$. Suppose $\gamma = \alpha_1$. Then (since $\alpha_1^3 = 1$), $F_{B_1^t B_1^{-1}}[\alpha_1] = \alpha_1^{k+1} + 1 + Q_{k-1}[\alpha_1] = \alpha_1^{k+1} + \alpha_1^{k-1} + \alpha_1^{k-2} + \cdots + \alpha_1$. Recall that $\alpha_1^2 + \alpha_1 + 1 = 0$. Thus, if $k - 1 \equiv 0 \pmod 3$, $F_{B_1^t B_1^{-1}}[\alpha_1] = \alpha_1^2 + 0 = \alpha_1^2$, if $k - 1 \equiv 1 \pmod 3$, then $F_{B_1^t B_1^{-1}}[\alpha_1] = 1 + \alpha_1 = \alpha_1^2$, and if $k - 1 \equiv 2 \pmod 3$, $F_{B_1^t B_1^{-1}}[\alpha_1] = \alpha_1 + \alpha_1^2 + \alpha_1 = \alpha_1^2$. Suppose $\gamma = \alpha_1^2$. Then, $F_{B_1^t B_1^{-1}}[\alpha_1^2] = \alpha_1^{2k+2} + 1 + \alpha_1 Q_{k-1}[\alpha_1^2]$. Note that if $k \equiv 0 \pmod 3$, $Q_{k-1}[\alpha_1^2] = 0$, if $k \equiv 1 \pmod 3$, $Q_{k-1}[\alpha_1^2] = 1$ and if $k \equiv 2 \pmod 3$, $Q_{k-1}[\alpha_1^2] = \alpha_1$. Thus, if $k \equiv 0 \pmod 3$, then $F_{B_1^t B_1^{-1}}[\alpha_1^2] = \alpha_1^2 + 1 + \alpha_1 \cdot 0 = \alpha_1$, if $k \equiv 1 \pmod 3$, $F_{B_1^t B_1^{-1}}[\alpha_1^2] = \alpha_1 + 1 + \alpha \cdot 1 = 1$ and if $k \equiv 2 \pmod 3$, $F_{B_1^t B_1^{-1}}[\alpha_1^2] = 1 + 1 + \alpha_1 \cdot \alpha_1 = \alpha_1^2$. This shows first part of (5). The second part of (5) follows from (1), just choose $\alpha_1 \in \mathbb{F} - \mathbb{K}$ so that $F_{B_1^t B_1}[\gamma] \neq 0$.

Corollary 5.14. (1) *Let* $\alpha_1 \in \Xi$ *and let* $\theta \in \mathbb{F}$ *such that* $\theta + \theta^q = \alpha_1^{q+1}$*. Then we can pick* $\alpha, \beta \in \{\alpha_1, \alpha_1^q\}$ *such that if we set* $X = X(\alpha, \theta)$ *and* $Y = X(\beta, \theta)$*, then for* $\{T, Z\} = \{X, Y\}$ *and* $S \in \{TZ^t, T^{-1}Z^t, T^t, Z^t\}$*,* $\langle \mathcal{O}(v_1, S)\rangle = V$*. Further, if* $q = 2$*,* $\alpha = \beta$*.*

(2) *Suppose* $q = 4$ *and let* $\theta \in \mathbb{F}^*$*. Suppose* $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{F}^*$ *are distinct elements such that* $\theta + \theta^q = \alpha_i^{q+1}$*,* $1 \leq i \leq 3$*. Then there exist* $\beta \in \{\alpha_1, \alpha_2, \alpha_3\}$ *such that for* $X = X(\beta, \theta)$*, and* $S \in \{XX^t, X^{-1}X^t, X^t\}$*,* $\langle \mathcal{O}(v_1, S)\rangle = V$*.*

(3) *If* $q \neq 3$ *is odd, or* $q = 3$ *and* $k \not\equiv 1 \pmod 3$*, then there are* $\alpha, \beta \in \mathbb{K}^*$*, such that if we set* $X = X(\alpha, \theta)$ *and* $Y = X(\beta, \rho)$*, with* $\theta = \frac{1}{2}\alpha^2$ *and*

$\rho = \frac{1}{2}\beta^2$, *then for* $\{T, Z\} = \{X, Y\}$ *and* $S \in \{TZ^t, T^{-1}Z^t, T^t, Z^t\}$, $\langle \mathcal{O}(v_1, S)\rangle = V$.

(4) *If* $q = 3$ *and* $k \geq 4$, *let* $a = a_k(1, 1, -1, 1, 1, \dots, 1)$ *and* $b = b_k(1, 1, -1, 1, 1, \dots, 1)$. *Let* $X = \operatorname{diag}(a_k, 1, b_k^{-1})u(1, \frac{1}{2})$ *and* $Y = \operatorname{diag}(a, 1, b^{-1})u(1, \frac{1}{2})$. *Then for* $\{T, Z\} = \{X, Y\}$ *and* $S \in \{TZ^t, T^{-1}Z^t, T^t, Z^t\}$, $\langle \mathcal{O}(v_1, S)\rangle = V$.

Proof. For (1), pick $\alpha, \beta \in \{\alpha_1, \alpha_1^q\}$. Let $B = u_1^{k+1}(-\alpha^q)\operatorname{diag}(1, b_k)$ and $B_1 = u_1^{k+1}(-\beta^q)\operatorname{diag}(1, b_k)$. By 5.11.8, for $\epsilon \in \{1, -1\}$, $\langle \mathcal{O}(v_1, X^\epsilon Y^t)\rangle = V$, iff $-\theta^{1-q}$ is not a root of $F_{B_1^t B^\epsilon}$. Note that since $\theta + \theta^q = \alpha_1^{q+1} \neq 0$, $\theta^{1-q} \neq -1$. Hence, using 5.13.3 (when $q > 2$, notice that $\alpha_1^2 \notin \mathbb{K}$ follows from the equation $\alpha_1^q + \alpha_1 = \alpha_1^{q+1}$), or 5.13.5 (when $q = 2$), we can pick $\alpha, \beta \in \{\alpha_1, \alpha_1^q\}$ such that $-\theta^{1-q}$ is not a root of $F_{B^t B_1^\epsilon}$ and not a root of $F_{B_1^t B^\epsilon}$ (with $\alpha = \beta$ when $q = 2$, by 5.13.5). Of course, by 5.11.9, $\langle \mathcal{O}(v_1, Y^t)\rangle = V = \langle \mathcal{O}(v_1, X^t)\rangle$, this shows (1).

The proof of (2) is similar. Setting $X_i = X(\alpha_i, \theta)$ and

$$B_i = u_1^{k+1}(-\alpha_i^q)\operatorname{diag}(1, b_k), \qquad 1 \leq i \leq 3,$$

we see, using 5.11.8, that for $\epsilon \in \{1, -1\}$, $\langle \mathcal{O}(v_1, X_i^\epsilon X_i^t)\rangle = V$, iff $-\theta^{1-q}$ is not a root of $F_{B_i^t B_i^\epsilon}$. Again we observe that $\theta^{1-q} \neq -1$. Further, since α_1, α_2 and α_3 are distinct, also, $\alpha_1^{2q}, \alpha_2^{2q}$ and α_3^{2q} are distinct, so by 5.13.1, there exists $1 \leq i \leq 3$, such that $\gamma = -\theta^{1-q}$ is not a root of the polynomial $F_{B_i^t B_i}$ and $F_{B_i^t B_i^{-1}}$.

For (3), notice first that, by 5.3.5, given $\alpha \in \mathbb{K}^*$, if we set $\theta = \theta(\alpha) = \frac{1}{2}\alpha^2$, then $X(\alpha, \theta) \in L$ and $\theta^{q-1} = 1$. Hence if $q > 3$, (3) follows from 5.11.8 and 5.13.4 (in the same way as we proved (1) and (2), noticing that since $\theta \in \mathbb{K}^*$, $\theta^{1-q} = 1$), and if $q = 3$, take $\alpha = \beta = 1$ and use 5.6.1. Finally (4) follows similarly using 5.11.8 and 5.6.2.

Theorem 5.15. (1) *We can pick* $\theta, \alpha, \beta \in \mathbb{F}$, *with* $\theta + \theta^q = \alpha^{q+1} = \beta^{q+1}$, *such that if we set* $X = X(\alpha, \theta)$ *and* $Y = X(\beta, \theta)$, *then:*

(i) *For* $\{T, Z\} = \{X, Y\}$ *and* $S \in \{TZ^t, T^{-1}Z^t, T^t, Z^t\}$, $\langle \mathcal{O}(v_1, S)\rangle = V$ *and:*

(ii) *For* $S \in \{TZ^t, T^{-1}Z^t\}$, $\alpha(F_S, 1) \neq 0$.

(2) *The commuting graph* $\Delta(L')$ *is balanced.*

Proof. For (1), suppose first that $q \neq 2, 4$. Then, by 5.10.10iii, we can find $\alpha_1 \in \Xi$, and $\theta \in \mathbb{F} - \mathbb{K}$, with $\theta + \theta^q = \alpha_1^{q+1}$, such that *for all* $\alpha, \beta \in \{\alpha_1, \alpha_1^q\}$, if we set $X = X(\alpha, \theta)$ and $Y = X(\beta, \theta)$, $\alpha(F_{T^\epsilon Z^t}, 1) \neq 0$, for $T, Z \in \{X, Y\}$ and $\epsilon \in \{1, -1\}$. Now, use 5.14.1, to pick $\alpha, \beta \in \{\alpha_1, \alpha_1^q\}$, such that for $\{T, Z\} = \{X, Y\}$ and $S \in \{TZ^t, T^{-1}Z^t, T^t, Z^t\}$, $\langle \mathcal{O}(v_1, S)\rangle = V$. This shows (1), in case $q \neq 2, 4$.

Suppose next that $q = 2$. Let $\alpha \in \mathbb{F} - \mathbb{K}$. Then, by 5.10.10i, for $X_1 \in \{X(\alpha,\alpha), X(\alpha^q,\alpha)\}$, and $\epsilon \in \{1,-1\}$, $\alpha(F_{X_1^\epsilon X_1^t}, 1) \neq 0$. By 5.14.1, there exists $X \in \{X(\alpha,\alpha), X(\alpha^q,\alpha)\}$, such that $V = \langle \mathcal{O}(v_1, S)\rangle$, for $S \in \{XX^t, X^{-1}X^t, X^t\}$, so (1) holds in case $q = 2$, choosing $Y = X$. The proof of (1) in case $q = 4$, is similar,using 5.10.10ii and 5.14.2.

We proceed with the proof of (2). Set $\Lambda = \Delta(L)$. Suppose $L \simeq SU(n,q)$ and let $X, Y \in L$ be as in (1). We show that $B_\Lambda(X, Y^t)$ holds. The proof that $B_\Lambda(Y, X^t)$ holds is symmetric and by 1.9, Λ is balanced. Let $S \in \{XY^t, X^{-1}Y^t, Y^t\}$. Suppose $R \in \Lambda^{\leq 2}(X) \cap \Lambda^{\leq 1}(S)$. By 5.12,

$$(*) \qquad v_1 \text{ is a characteristic vector of } R.$$

Now if $S = Y^t$, then S commutes with R, so since $V = \langle \mathcal{O}(v_1, Y^t)\rangle$, $(*)$ implies that $R \in Z(L)$, a contradiction. Suppose $S \in \{XY^t, X^{-1}Y^t\}$. Then, by (ii) of (1), $\gcd\{\{i : \alpha(F_S, i) \neq 0\} \cup \{n\}\} = 1$, so, by $(*)$ and 1.14.5, $R \in Z(L)$, a contradiction.

Suppose $L \simeq SO_n(q)$. Pick X, Y as in 5.14.3 and 5.14.4. Since $Z(L) = 1$, to show $B_\Lambda(X, Y^t)$ holds, it suffices, by 1.14.1, to show that $V = \langle \mathcal{O}(v_1, S)\rangle$, for $S \in \{XY^t, X^{-1}Y^t, Y^t\}$, but this holds by the choice of X, Y. By symmetry also $B_\Lambda(Y, X^t)$ holds and the proof of the theorem is complete.

6. The Orthogonal Groups in odd characteristic and even dimension.

In this section $\mathbb{F}$ is a field of odd order and $n = 2k \geq 8$ is even. Let J be the following $n \times n$ matrix:

$$J = \begin{bmatrix} 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & \bar{1} & 0 & 0 \\ 0 & 0 & \cdot & \cdot & 0 & 1 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \bar{1} & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \nu \end{bmatrix}.$$

Let $L \simeq SO^\epsilon(\mathbb{F})$ be the subgroup of $SL_n(\mathbb{F})$ defined by $L = \{x \in SL_n(\mathbb{F}) : xJx^t = J\}$. Of course, for a suitable choice of ν ($\nu = (-1)^k$) $\epsilon = +$ and for a suitable choice of ν ($(-1)^k\nu$ a nonsquare in $\mathbb{F}$) $\epsilon = -$.

We continue with the notation of Section 1. In addition we let $f : V \times V \to \mathbb{F}$ be a bilinear form whose matrix with respect to the basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ is J.

6.1. *Let $u \in GL_n(q)$ be a matrix of the form*

$$u = \begin{bmatrix} 1 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 & 0 \\ \alpha_2 & 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ * & \alpha_3 & 1 & 0 & 0 & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ * & \cdot & \cdot & \cdot & * & \alpha_{n-2} & 1 & 0 & 0 \\ * & \cdot & \cdot & \cdot & \cdot & * & \alpha_{n-1} & 1 & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 \end{bmatrix} \qquad \alpha_i \in \mathbb{F}^*, \quad \text{for all } i.$$

Let $h \in GL(n, \mathbb{F}) - Z(GL(n, \mathbb{F}))$ be a matrix commuting with u. Then:

(1) *h has the form*

$$h = \begin{bmatrix} M & E \\ F & c \end{bmatrix}$$

where M is an $(n-1)\times(n-1)$ matrix commuting with $M_{n,n}(u)$, $c \in \mathbb{F}^*$, E is a column $(n-1)\times 1$ matrix of the form $(0, 0, \dots, \rho)^t$, F is a row $1 \times (n-1)$ matrix of the form $(\theta, 0, \dots, 0)$.

(2) *Suppose $u, h \in L$, and let $\rho, \theta \in \mathbb{F}$ as in (1). Then there exists $\epsilon \in \{1, -1\}$ such that $h_{i,i} = \epsilon$, for all $1 \le i \le n$. Further, $\theta = -\rho f(v_n, v_n)$.*

(3) *If $u, h \in L$, then there exists $\epsilon \in \{1, -1\}$, and $1 \le r' < n-1$, such that*

$$h - \epsilon I_n = \begin{bmatrix} t' & E \\ F & 0 \end{bmatrix}$$

with $t' \in \mathcal{T}_{n-1}(r')$ (see notation in 1.1.10).

(4) *Suppose $u, h \in L$ and let t' and r' be as in (3) and ρ as in (2). Suppose that either $\rho = 0$, or $r' \neq k - 1$. There exists $\epsilon \in \{1, -1\}$, $i \in \{1, 2\}$ and $1 \le r < n-1$, such that*

$$(h - \epsilon I_n)^i = \begin{bmatrix} t & 0_{n-1,1} \\ 0_{1,n-1} & 0 \end{bmatrix}$$

where $t \in \mathcal{T}_{n-1}(r)$.

(5) *Suppose $u, h \in L$ and let t' and r' be as in (3) and ρ, θ as in (2). Suppose $r' = k - 1$ and $\rho \neq 0$. Then:*

(5i) *k is even.*

(5ii) *If, in addition, $(h-\epsilon I_n)^2 = 0$, then we may assume that $f(v_n, v_n) = 1$ (so $\nu = 1$) and if we set $d = t'_{k,1}$, then $d^2 = \theta^2$.*

Proof. Note that h commutes with the matrix $u - I_n$, and clearly for $1 \le i \le n-1$, $\operatorname{im}(u - I_n)^i = \mathcal{V}_{n-i-1}$. Since h commutes with $(u - I_n)^i$, h fixes $\operatorname{im}(u - I_n)^i$. Thus h fixes $\mathcal{V}_i$, for $1 \le i \le n-2$. Also $\ker(u - I_n) = \langle v_1, v_n \rangle$, so h fixes $\langle v_1, v_n\rangle$, thus h has the form

$$h = \begin{bmatrix} M & E \\ F & c \end{bmatrix}$$

with M some $(n-1)\times(n-1)$ matrix and E, F and c as in (1). Let $u_1 = M_{n,n}(u)$. Then

$$hu = \begin{bmatrix} Mu_1 & E \\ F & c \end{bmatrix} \quad \text{and} \quad uh = \begin{bmatrix} u_1M & E \\ F & c \end{bmatrix}$$

this shows (1).

For (2), note that $v_nh = \theta v_1 + cv_n$, thus $0 \neq f(v_n, v_n) = f(v_nh, v_nh) = c^2f(v_n, v_n)$. Thus $c = \epsilon$, for some $\epsilon \in \{1, -1\}$. Also, since u_1 commutes with M, 1.13.2 implies that there exists $\beta \in \mathbb{F}$, such that $h_{i,i} = \beta$, for all $1 \leq i \leq n-1$. Since v_k is a nonsingular vector, it is easy to check that we must have $\beta = 1$ or -1. Since $\det(h) = 1$, $\beta = \epsilon$ and the first part of (2) is proved. For the second part we have $0 = f(v_{n-1}, v_n) = f(v_{n-1}h, \theta v_1 + \epsilon v_n) = f(v' + \epsilon v_{n-1} + \rho v_n, \theta v_1 + \epsilon v_n)$, with $v' \in \mathcal{V}_{n-2}$. But $f(v_1, v') = f(v_n, v') = 0$. Thus $0 = f(v_{n-1}, v_n) = f(\epsilon v_{n-1} + \rho v_n, \theta v_1 + \epsilon v_n) = \epsilon\theta + \epsilon\rho f(v_n, v_n)$ and the second part of (2) is proved.

Next note that by (1), $u_1 := M_{n,n}(u)$, commutes with M so, by 1.13 and (2), $(M - \epsilon I_{n-1}) \in \mathcal{T}_{n-1}(r')$, for some $1 \leq r' < n-1$. Thus (3) follows from (1) and (2).

For (4), we use (3). If $\rho = 0$, then, by (2) also $\theta = 0$, and so by (3), (4) holds with $i = 1$, $r = r'$ and $t = t'$. Suppose $\rho \neq 0$. Note that EF is an $(n-1)\times(n-1)$ matrix whose $(n-1, 1)$-entry is $\rho\theta$ and for $(i, j) \neq (n-1, 1)$, $(EF)_{ij} = 0$. Further $t'E = 0_{n-1,1}$ (the last column of t' is zero), $Ft' = 0_{1,n-1}$ (the first row of t' is zero) and $FE = 0$. Thus

$$(h - \epsilon I_n)^2 = \begin{bmatrix} t' & E \\ F & 0 \end{bmatrix} \cdot \begin{bmatrix} t' & E \\ F & 0 \end{bmatrix} = \begin{bmatrix} (t')^2 + EF & 0_{n-1,1} \\ 0_{1,n-1} & 0 \end{bmatrix}.$$

Since we are assuming that $\rho \neq 0$ and $r' \neq k-1$, either $r' > k-1$, in which case $(t')^2 = 0$, and $t = EF \in \mathcal{T}_{n-1}(n-2)$. Or $r' < k-1$, in which case, $(t')^2 \in \mathcal{T}_{n-1}(r)$, for some $1 < r < n-2$, and then $t := (t')^2 + EF \in \mathcal{T}_{n-1}(r)$. This shows (4).

Finally assume the hypotheses of (5). Suppose first that k is odd. Let $j = \frac{k+1}{2}$, then $r' + j = (k-1) + \frac{k+1}{2} = \frac{3k-1}{2}$ and $t'_{r'+j,j} \neq 0$. But $v_{r'+j}h = v' + t'_{r'+j,j}v_j + \epsilon v_{r'+j}$, with $v' \in \mathcal{V}_{j-1}$. But $0 = f(v_{r'+j}, v_{r'+j}) = f(v_{r'+j}h, v_{r'+j}h) = 2\epsilon t'_{r'+j}f(v_j, v_{r'+j}) \neq 0$, a contradiction. Hence k is even. To prove (5ii), set $d = t'_{k,1}$. We claim that $t'_{n-1,k} = d$. Indeed, $0 = f(v_k, v_{n-1}) = f(v_kh, v_{n-1}h) = f(dv_1 + \epsilon v_k, t'_{n-1,1}v_1 + \cdots + t'_{n-1,k}v_k + \epsilon v_{n-1} + \rho v_n) = \epsilon d + (-1)^{k+1}\epsilon t'_{n-1,k}$, thus $t'_{n-1,k} = (-1)^k d = d$. Also the $(n-1, 1)$-entry of $(t')^2$ is d^2 and the remaining entries of $(t')^2$ are zero. Since $(h - \epsilon I_n)^2 = 0$, we must have (see the proof of (4)), $(t')^2 + EF = 0$, so

$d^2 + \theta\rho = 0$. But $\theta\rho = -\rho^2 f(v_n, v_n)$ (see (2)), so $d^2 = \rho^2 f(v_n, v_n)$. Hence, $f(v_n, v_n)$ is a square in $\mathbb{F}$, so we may take $f(v_n, v_n) = 1$. Then $d^2 = \rho^2$, and since, by (2), $\theta = -\rho$, $d^2 = \theta^2$.

Notation. For the remainder of this section, we fix the following notation. Let $\beta_1, \dots, \beta_{k-2}, \gamma_1, \dots, \gamma_{k-2} \in \mathbb{F}^*$. Let also $\alpha, \beta \in \mathbb{F}^*$. We set $a = a_{k-1}(\beta_1, \dots, \beta_{k-2})$, $a_1 = a_{k-1}(\gamma_1, \dots, \gamma_{k-2})$, $b = b_{k-1}(\beta_1, \dots, \beta_{k-2})$, $b_1 = b_{k-1}(\gamma_1, \dots, \gamma_{k-2})$, $g = \operatorname{diag}(a, 1, b^{-1})$, $g_1 = \operatorname{diag}(a_1, 1, b_1^{-1})$, $B = b_k(\alpha, \beta_1, \dots, \beta_{k-2})$, $B_1 = b_k(\beta, \gamma_1, \dots, \gamma_{k-2})$, $u = u^{n-1}(\alpha, \frac{1}{2}\alpha^2)$, $w = u^{n-1}(\beta, \frac{1}{2}\beta^2)$ (notation as in 5.4.1), $\mathcal{X} = gu$ and $\mathcal{Y} = g_1 w$. Finally, we let $X = \operatorname{diag}(\mathcal{X}, 1)$ and $Y = \operatorname{diag}(\mathcal{Y}, 1)$.

6.2. *Let $\epsilon' \in \{1, -1\}$, and $\mathcal{S} \in \{\mathcal{Y}^t, \mathcal{X}^{\epsilon'}\mathcal{Y}^t\}$. Set $S = \operatorname{diag}(\mathcal{S}, 1)$ and let $R \in \Delta^{\le 2}(X) \cap \Delta^{\le 1}(S)$. Then v_1 is characteristic vector of R.*

Proof. Let $h \in \Delta^{\le 1}(X) \cap \Delta^{\le 1}(R)$. Note that by 5.11.3, X satisfies the hypothesis for u in 6.1, there exists $\epsilon \in \{-1, 1\}$ and $1 \le r' < n-1$, such that

$$h - \epsilon I_n = \begin{bmatrix} t' & E \\ F & 0 \end{bmatrix}$$

with $t' \in \mathcal{T}_{n-1}(r')$.

We'll show that there exists $i \in \{1, 2\}$ such that if we set $T := (h - \epsilon I_n)^i$, then T, S and R satisfy all the hypotheses of 1.15, for a suitable choice of j, m and ℓ. Then the lemma follows from 1.15. First, $R^{-1}TR = T$ and $[R, S] \in Z(L)$, so hypothesis (e) of 1.15 is satisfied. Note next that by 5.11.5:

(i) S satisfies hypothesis (a) of 1.15 for any $j \le k-2$.

We now distinguish two cases as follows.

Case 1. There exists $i \in \{1, 2\}$ and $1 \le r < n-1$, such that

$$(h - \epsilon I_n)^i = \begin{bmatrix} t & 0_{n-1,1} \\ 0_{1,n-1} & 0 \end{bmatrix} \quad \text{where } t \in \mathcal{T}_{n-1}(r).$$

Let $T := (h - \epsilon I_n)^i$, with i as above. Observe that $M_{n,n}(ST) = \mathcal{S}t$, hence we get from 5.11.6 (replacing k by $k-1$) that:

(ii) If $r \ge k-1$, then $v_{k-1} \notin \ker(ST)$ and $\mathcal{V}_{k-2} \subseteq \ker(ST)$.

Next observe that if $r > k-1$ (and (ii) necessarily holds), $n-r-1 \le k-1$ and $\operatorname{im}(T) = \mathcal{V}_{n-r-1}$ is R-invariant. Thus:

(iii) If $r > k-1$, then $n-r-1 \le k-1$ and $\mathcal{V}_{n-r-1}$ is R-invariant.

Hence if $r > k-1$, take $j = k-2$, $m = n-r-1$ and $\ell = 1$ and, by (i), (ii) and (iii), all hypotheses of 1.15 are met, so we are done.

Next observe that if $r \leq k-1$, then $\ker(T) = \langle v_1, \dots, v_r, v_n\rangle$ and the radical of the form f, reduced to $\ker(T)$ is $\mathcal{V}_r$. Thus:

(iv) If $r \leq k-1$, then $\mathcal{V}_r$ is an R-invariant subspace and $v_{r+1} \notin \ker(T)$.

Thus if $r = k-1$, take $m = r$, $j = k-2$ and $\ell = 1$, and, by (i), (ii) and (iv) we are done, while if $r < k-1$, take $j = m = r$ and $\ell = 0$ and observe that by (i) and (iv) we are done.

Case 2. $r' = k-1$, $\rho \neq 0 \neq \theta$, $\nu = f(v_n, v_n) = 1$ and for $d = t'_{k,1}$, $d^2 = \theta^2$.

Note that by 6.1.4 and 6.1.5, either Case 1 holds or Case 2 holds. Let $T = X - \epsilon I_n$. Write $d = -\epsilon''\theta$, with $\epsilon'' \in \{1, -1\}$. Observe that $\ker(T) = \{v_1, \dots, v_{k-1}, v_n + \epsilon'' v_k\}$. First we claim that:

(v) There exists $v \in \mathcal{V}_{k-1}$ such that modulo $\mathcal{V}_{k-1}$, we have

$$v_k S^{-1} \equiv \eta v_k + \mu v_{k+1}, \quad \text{with } \eta \in \mathbb{F} \text{ and } \mu \in \mathbb{F}^*$$
$$\{(v_n + \epsilon'' v_k) - \epsilon'' v\} S^{-1} \equiv v_n - \epsilon'' v_k.$$

Indeed, we use 5.11.7. We take in 5.11, $n = 2k-1 = 2(k-1)+1$, α, β, and ρ (of 5.11) in the fixed field of σ_q (so $\rho^{1-q} = 1$). Thus, for all possibilities of $\mathcal{S}$ the following holds:

(vi) There exists $v \in \mathcal{V}_{k-1}$, $\eta \in \mathbb{F}$ and $\mu \in \mathbb{F}^*$ such that

$$v_k \mathcal{S}^{-1} \equiv \eta v_k + \mu v_{k+1} \pmod{\mathcal{V}_{k-1}}$$
$$v \mathcal{S}^{-1} \equiv (\eta + 1) v_k + \mu v_{k+1} \pmod{\mathcal{V}_{k-1}}.$$

Where in all cases $\mu = -\beta$. If $\mathcal{S} = \mathcal{Y}^t$, $\eta = 1$, while

if $\mathcal{S} = \mathcal{X}^{\epsilon'} \mathcal{Y}^t$, $\eta = 1 + \epsilon' \alpha\beta$.

Thus, by (vi), modulo $\mathcal{V}_{k-1}$ we get that

$$\begin{aligned}
&\{(v_n + \epsilon'' v_k) - \epsilon'' v\} S^{-1} \\
&\equiv v_n + \epsilon''\{\eta v_k + \mu v_{k+1}\} - \epsilon''\{(\eta+1) v_k + \mu v_{k+1}\} \\
&\equiv v_n + \{\eta\epsilon'' - (\eta+1)\epsilon''\} v_k + (\mu\epsilon'' - \mu\epsilon'') v_{k+1} \\
&\equiv v_n - \epsilon'' v_k.
\end{aligned}$$

This shows (v).

Let v and ϵ'' be as in (v). Since $v, v_n + \epsilon'' v_k \in \ker(T)$, $\mathcal{U} := \langle v S^{-1}, (v_n + \epsilon'' v_k) S^{-1}\rangle \subseteq \ker(ST)$. Notice that (v) implies that $v_n - \epsilon'' v_k \in \mathcal{U} + \mathcal{V}_{k-1}$ and also that $v S^{-1} \equiv \mu v_{k+1} \pmod{\mathcal{V}_k}$ (μ as in (v)). Hence we conclude that $\mathcal{U} \cap \ker(T) = (0)$. Since $\dim(\mathcal{U}) = 2$, and since $\dim(\ker(T)) = k$, we get that $\dim(\ker(T) \cap \ker(ST)) \leq k-2$. But $\mathcal{V}_{k-2} \subseteq \ker(ST)$ and hence $\ker(T) \cap \ker(ST) = \mathcal{V}_{k-2}$. Clearly $\ker(T) \cap \ker(ST)$ is R-invariant, so we conclude that:

(vii) $\mathcal{V}_{k-2}$ is R-invariant.

Observe that (ii) holds here as well, since $M_{n,n}(ST) = \mathcal{S}t$, holds here as well. Hence if we take $m = k - 2 = j$ and $\ell = 1$, we see that all hypotheses of 1.15 hold here as well and the proof of 6.2 is complete.

6.3. *Let* $\epsilon \in \{1, -1\}$ *and let* $S \in \{Y^t, X^\epsilon Y^t\}$*. Set* $\mathcal{S} = M_{n,n}(S)$ *and suppose* $\langle \mathcal{O}(v_1, \mathcal{S})\rangle = \mathcal{V}_{n-1}$*. Then* $d_\Lambda(X, S) > 3$*, where* $\Lambda = \Delta(L)$*.*

Proof. Let $R \in \Delta^{\le 2}(X) \cap \Delta^{\le 1}(S)$. By 6.2, v_1 is a characteristic vector of R and since $\langle \mathcal{O}(v_1, \mathcal{S})\rangle = \mathcal{V}_{n-1}$, $\mathcal{V}_{n-1}$ is an R-invariant subspace. Thus $\mathcal{V}_{n-1}^\perp = \langle v_n\rangle$ is R-invariant as well. Set $R_1 = M_{n,n}(R)$. Since $[R, S] \in Z(L)$, $[R_1, \mathcal{S}] = \pm I_{n-1}$ and since $\det([R_1, \mathcal{S}]) = 1$, $[R_1, \mathcal{S}] = I_{n-1}$. Thus $[R, S] = 1$, and since v_1 is a characteristic vector of R and $\langle \mathcal{O}(v_1, \mathcal{S})\rangle = \mathcal{V}_{n-1}$, $R_1 = \pm I_{n-1}$. Of course $R_{n,n} \in \{1, -1\}$ and since $\det(R) = 1$, $R \in Z(L)$, a contradiction.

Theorem 6.4. $\Delta(L)$ *is balanced.*

Proof. In 5.14.3 and 5.14.4, we showed that we can pick $\mathcal{X}, \mathcal{Y}$ such that for $\{\mathcal{T}, \mathcal{Z}\} = \{\mathcal{X}, \mathcal{Y}\}$, $\epsilon \in \{1, -1\}$ and $\mathcal{S} \in \{\mathcal{T}^t, \mathcal{T}^\epsilon \mathcal{Z}^t\}$, $\langle \mathcal{O}(v_1, \mathcal{S})\rangle = \mathcal{V}_{n-1}$. Hence the theorem follows from 6.3 and by definition.

7. The Orthogonal Groups in even dimension and even characteristic.

In this section $n = 2k \ge 8$ is even and $\mathbb{F}$ is a field of even order. We keep the notation of Section 1. In particular V is a vector space of dimension n over $\mathbb{F}$ and $\mathcal{B} = \{v_1, \dots, v_n\}$ is our fixed basis of V. Let f be the symplectic form on V whose matrix with respect to $\mathcal{B}$ is

$$J = \begin{bmatrix} 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & 1 & 0 \\ 0 & 0 & \cdot & \cdot & 0 & 1 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix}.$$

For $\epsilon \in \{+, -\}$ let Q^ϵ be the quadratic form on V defined as follows. First $Q^\epsilon(v + w) = Q^\epsilon(v) + Q^\epsilon(w) + f(v, w)$, for all $v, w \in V$. Second, $Q^\epsilon(v_i) = 0$, for all $1 \le i \le k-1$ and all $k+2 \le i \le n$. We define $Q^\epsilon(v_k) = Q^\epsilon(v_{k+1}) = \nu_\epsilon$, where $\nu_\epsilon = 0$, when $\epsilon = +$ and when $\epsilon = -$, $\nu_\epsilon \ne 0$, is such that $\nu_\epsilon\lambda^2 + \lambda + \nu_\epsilon$ is an irreducible polynomial in $\mathbb{F}[\lambda]$. Of course V is an orthogonal space of type ϵ in the respective cases. We let $Q = Q^\epsilon$. We denote by $Q^\epsilon(V, Q)$ the full orthogonal group of type $\epsilon \in \{+, -\}$ in the respective cases. We let L be the commutator subgroup of $O^\epsilon(V, Q)$. Thus L is a simple group and L has index 2 in $O^\epsilon(V, Q)$. The purpose of this section is to prove

Theorem 1.6 for L. For that we'll show that L is closed under transpose (see 1.4.3) and indicate an element $X \in L$ such that $B_\Lambda(X, X^t)$ holds, where $\Lambda = \Delta(L)$. Then, by 1.9.2, Λ is balanced. We'll define X shortly. The following Theorem is useful.

7.1. *Let $g \in O^\epsilon(V,Q)$. Then $g \in L$ if and only if $\dim C_V(g)$ is even.*

Proof. See [**3**], Theorem 3.

7.2. *L is closed under transpose.*

Proof. Regard J above as an element of $GL(V)$. Then J is an involution and $J^t = J$ (J is symmetric). We claim that $J \in Q^\epsilon(V,Q)$. Indeed $JJJ^t = J \in O(V,f)$ and since $v_iJ = v_{n+1-i}$, for all $1 \le i \le n$, J preserves the quadratic form Q, since in both types $Q(v_i) = Q(v_{n+1-i})$. But for $g \in L$, $g^t = Jg^{-1}J$, so $g^t \in L$.

Notation 7.3. (1) Let $g \in GL(V)$ such that $g = \operatorname{diag}(I_{k-2}, s, I_{k-2})$, where s is some 4×4 matrix. We denote s by $s(g)$.

(2) Throughtout this section $u := \operatorname{diag}(I_{k-2}, s, I_{k-2})$, where

$$s = s(u) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

(3) Throughout this section we let

$$g = \operatorname{diag}(a_k, b_k^{-1})$$
$$X = gu$$

where for $m \ge 1$, a_m and b_m are as in 1.1.9. Note that since $\operatorname{char}(\mathbb{F}) = 2$, $a_m = b_m$.

(4) We denote by $\mathcal{C}$, the ordered basis $(w_1 \dots, w_n)$, where $w_i = v_i$, for $1 \le i \le k-2$, $w_{k-1} = v_{k-1} + v_k + v_{k+1}$, $w_i = v_{i+2}$, for $k \le i \le n-2$, $w_{n-1} = v_k + v_{k+1}$ and $w_n = v_k + v_{k+2}$. Thus

$$\mathcal{C} = (v_1, v_2, \dots, v_{k-2}, v_{k-1} + v_k + v_{k+1}, v_{k+2}, \dots, v_n, v_k + v_{k+1}, v_k + v_{k+2}).$$

7.4. (1)

$$s(u) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \qquad (s(u))^t = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

(2) $u^{-1} = u$.

(3)

$$s(uu^t) = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

(4)

$$s((uu^t)^{-1}) = s(u^t u) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

(5) $s(u^{-1}u^t) = s(uu^t)$ *and* $s((u^{-1}u^t)^{-1}) = s(u^t u)$.
(6) $[g^t, u] = 1$.

Proof. (1) is by definition. Clearly $u^{-1} = u$. For (3) and (4), we compute

$$s(uu^t) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

$$s((uu^t)^{-1}) = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

(5) follows from (2). For (6) we have, $v_i g^t u = v_i g^t = v_i u g^t$, for $i \notin \{k+1, k+2\}$. $v_{k+1} g^t u = (v_{k+1} + \cdots + v_n)u = v_{k-1} + v_k + \cdots + v_n$ and $v_{k+1} u g^t = (v_{k-1} + v_{k+1}) g^t = v_{k-1} + v_k + \cdots + v_n$. $v_{k+2} g^t u = (v_{k+2} + \cdots + v_n)u = v_k + v_{k+2} + \cdots + v_n$ and $v_{k+2} u g^t = (v_k + v_{k+2}) g^t = v_k + v_{k+2} + \cdots + v_n$.

7.5. (1)

$$X=\left[\begin{array}{cccccccccccccccc}
1&0&\cdot&\cdot&\cdot&\cdot&\cdot&0&&&&&&&&\\
1&1&0&\cdot&\cdot&\cdot&\cdot&0&&&&&&&&\\
0&1&1&0&\cdot&\cdot&\cdot&0&&&&&&&&\\
0&0&1&1&0&\cdot&\cdot&0&&&&&&&&\\
\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&&&&&&&&\\
\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&&&&&&&&\\
0&\cdot&\cdot&\cdot&0&1&1&0&&&&&&&&\\
0&\cdot&\cdot&\cdot&\cdot&0&1&1&&&&&&&&\\
&&&&&&1&0&1&0&\cdot&\cdot&\cdot&\cdot&\cdot&0\\
&&&&&&1&1&1&1&0&\cdot&\cdot&\cdot&\cdot&0\\
&&&&&&1&1&1&1&1&0&\cdot&\cdot&\cdot&0\\
&&&&&&1&1&1&1&1&1&0&\cdot&\cdot&0\\
&&&&&&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot\\
&&&&&&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot\\
&&&&&&1&1&1&\cdot&\cdot&\cdot&\cdot&1&1&0\\
&&&&&&1&1&1&\cdot&\cdot&\cdot&\cdot&1&1&1
\end{array}\right]$$

where the blank spots are zeros. Also the upper submatrix of X is a $k \times k$ matrix and the lower submatrix of X is a $k \times (k+2)$ matrix.

(2) *The matrix of X with respect to the basis $\mathcal{C}$ is*

$$[X]_{\mathcal{C}}=\left[\begin{array}{cccccccccccccccc}
1&0&\cdot&\cdot&\cdot&\cdot&0&&&&&&&&&\\
1&1&0&\cdot&\cdot&\cdot&0&&&&&&&&&\\
0&1&1&0&\cdot&\cdot&0&&&&&&&&&\\
0&0&1&1&0&\cdot&0&&&&&&&&&\\
\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&&&&&&&&&\\
\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&&&&&&&&&\\
0&\cdot&\cdot&0&1&1&0&&&&&&&&&\\
0&\cdot&\cdot&\cdot&0&1&1&&&&&&&&&\\
&&&&&&1&1&0&\cdot&\cdot&\cdot&\cdot&0&&\\
&&&&&&1&1&1&0&\cdot&\cdot&\cdot&0&&\\
&&&&&&1&1&1&1&0&\cdot&\cdot&0&&\\
&&&&&&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&&\\
&&&&&&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&&\\
&&&&&&1&1&\cdot&\cdot&\cdot&\cdot&1&0&&\\
&&&&&&1&1&\cdot&\cdot&\cdot&\cdot&\cdot&1&&\\
&&&&&&&&&&&&&&1&0\\
&&&&&&&&&&&&&&1&1
\end{array}\right]$$

where the blank spots are zeros. Also the upper submatrix of $[X]_{\mathcal{C}}$ is a $(k-1) \times (k-1)$ matrix, the middle submatrix of $[X]_{\mathcal{C}}$ is a $(k-1) \times k$ matrix and of course the lower submatrix of $[X]_{\mathcal{C}}$ is a 2×2 matrix.

(3) $X \in L$.

Proof. (1) and (2) are easy calculations and we omit the details. Next, since $\mathcal{V}_{k-1}$ and $\langle v_{k+2}, \dots, v_n\rangle$ are totally singular subspaces (in both types), $Q(v_iX) = 0$, for $1 \le i \le k-1$. Also, for $k+2 \le i \le n$, $Q(v_iX) = Q(v_{k-1} + v_k + v_{k+1} + v_{k+2} + \cdots + v_i) = Q(v_{k-1} + v_k + v_{k+1} + v_{k+2}) = Q(v_{k-1} + v_{k+2}) + Q(v_k + v_{k+1}) = 1 + 1 = 0$. Further, for $s \in \{k, k+1\}$, $Q(v_sX) = Q(v_{k-1} + v_s) = Q(v_s)$.

We leave it for the reader to verify that $XJX^t = J$, so $X \in O(V, f)$. Since $C_V(X) = \langle v_1, v_k + v_{k+1}\rangle$, $X \in L$, by 7.1.

7.6. *Let B be the following $(k+1) \times (k+1)$ matrix*

$$B = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & 1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & 1 & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 & 1 \end{bmatrix}.$$

Then:

(1)
$$B^{-1} = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & 1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 1 & 1 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & 0 \\ 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix}.$$

(2)
$$B^tB = \begin{bmatrix} 0 & 1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 1 & 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & 1 & 0 & 1 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 & 1 \end{bmatrix} \qquad B^tB^{-1} = \begin{bmatrix} 0 & 1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 \\ 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & 1 \end{bmatrix}.$$

Proof. (1) is easy to check. For (2), we compute

$$B^tB = \begin{bmatrix} 1 & 0 & 1 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & 1 & 1 \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & 1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & 1 & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 1 & 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & 1 & 0 & 1 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 & 1 \end{bmatrix},$$

$$B^tB^{-1} = \begin{bmatrix} 1 & 0 & 1 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & 1 & 1 \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & 1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 1 & 1 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & 0 \\ 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix}$$

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

7.7. *Set* $a = a_{k-1}$ *and* $v = v_k + v_{k+1}$. *Let* B *be as in* 7.6 *and let* $\epsilon \in \{-1, 1\}$. *Then:*

(1) $XX^t = guu^tg^t$, $(XX^t)^{-1} = (g^t)^{-1}(u^tu)g^{-1}$.
(2) $X^{-1}X^t = uu^tg^{-1}g^t$ *and* $(X^{-1}X^t)^{-1} = (g^t)^{-1}gu^tu$.
(3) $X = \begin{bmatrix} a & 0_{k-1,k+1} \\ E & B^{-1} \end{bmatrix}$ *with* E *some* $(k+1) \times (k-1)$ *matrix.*

(4) $X^\epsilon X^t = \begin{bmatrix} a^\epsilon a^t & R_{1,2} \\ R_{2,1} & R_{2,2} \end{bmatrix}$ $(X^\epsilon X^t)^{-1} = \begin{bmatrix} R'_{1,1} & R'_{1,2} \\ R'_{2,1} & B^t B^\epsilon \end{bmatrix}$ *with* $R'_{1,1}, R_{2,2}, R_{1,2}, R'_{1,2}, R_{2,1}, R'_{2,1}$ *some* $(k-1)\times(k-1)$, $(k+1)\times(k+1)$, $(k-1)\times(k+1)$, $(k-1)\times(k+1)$, $(k+1)\times(k-1)$, $(k+1)\times(k-1)$ *matrices respectively. Further, the first* $k-2$ *rows of* $R_{1,2}$ *are zero.*

(5) *Let* $S \in \{X^t, X^\epsilon X^t\}$. *Then for* $1 \le i \le k-2$, $v_i S = w + v_{i+1}$, *with* $w \in \mathcal{V}_i$. *In particular,* $\mathcal{V}_{k-1} \subseteq \langle \mathcal{O}(v_1, S)\rangle$.

(6) *Let* $S \in \{X^t, X^\epsilon X^t\}$. *Then* $v_{k-1}S = w + v_n$, *with* $w \in \mathcal{V}_{n-1}$.

(7)(7i) *Let* $S = X^t$, *then* $v_{k-1}S^{-1} = v_{k-1} + v_k + v_{k+1} + v_{k+2}$, $v_k S^{-1} = v_k + v_{k+2}$, *and* $v_{k+1}S^{-1} = v_{k+1} + v_{k+2}$.

(7ii) *Let* $S = XX^t$, *then* $v_{k-1}S^{-1} = v_{k+2}$, $v_k S^{-1} = v_{k+1} + v_{k+2}$, *and* $v_{k+1}S^{-1} = v_k + v_{k+2}$.

(7iii) *Let* $S = X^{-1}X^t$, *then* $v_{k-1}S^{-1} = v_{k-2} + v_{k+2}$, $v_k S^{-1} = v_{k+1} + v_{k+2}$, *and* $v_{k+1}S^{-1} = v_k + v_{k+2}$.

(8) $\langle \mathcal{O}(v_1, X^t)\rangle = \langle \mathcal{V}_{k-1}, v + v_{k+2}, v_{k+3}, \ldots, v_n\rangle$. *Further if we set* $\mathcal{W} = \langle \mathcal{O}(v_1, X^t)\rangle$, *then* $\mathcal{W}^\perp = \langle v, v_{k-1} + v_k\rangle$, $vX^t = v$ *and* $(v_{k-1} + v_k)X^t = v + (v_{k-1} + v_k)$.

(9) *Let* $S = XX^t$. *Then:*

(9i) *If* $k \equiv 1$ *or* $2 \pmod 3$, *then*

$$\langle \mathcal{O}(v_1, S)\rangle = \langle \mathcal{V}_{k-1}, v, v_{k+2}, v_{k+3}, \ldots, v_n\rangle.$$

(9ii) *If* $k \equiv 0 \pmod 3$, *then*

$$\langle \mathcal{O}(v_1, S)\rangle = \Big\langle \mathcal{V}_{k-1}, v_{k+2}, v + v_{k+3j}, v + v_{k+3j+1}, v_{k+3j+2}, v + v_n : 1 \le j \le \frac{1}{3}k - 1\Big\rangle.$$

Further, in (9ii), *if we set* $\mathcal{W} = \langle \mathcal{O}(v_1, S)\rangle$, *then* $\mathcal{W}^\perp = \langle v, v'\rangle$, *where*

$$v' = (v_1 + v_3) + (v_4 + v_6) + (v_7 + v_9) + \cdots + (v_{k-2} + v_k),$$

$vS = v$ *and* $v'S = v + v'$.

(10) *Let* $S = X^{-1}X^t$. *Then*

$$\langle \mathcal{O}(v_1, S)\rangle = \langle \mathcal{V}_{k-1}, v, v_{k+2}, v_{k+3}, \ldots, v_n\rangle.$$

Proof. (1) is obvious, recalling (see 7.4.2) that $u^{-1} = u$. For (2), we have $X^{-1}X^t = u^{-1}g^{-1}u^t g^t$. By 7.4.6, $[g^{-1}, u^t] = 1$, and (2) follows. For (3), just observe that X is given in 7.5.

(4) follows from (3), except that we must show that the first $k-2$ rows of $R_{1,2}$ are zero. This will of course follow from (5). To show (5), let $1 \le i \le k-2$. Suppose first that $S = X^t$. Then $v_i S = v_i u^t g^t = v_i g^t = v_i + v_{i+1}$. Next $v_i u u^t = v_i$, so $v_i X^{-1}X^t = v_i g^{-1} g^t$. Also $v_i g \in \mathcal{V}_i$, so $v_i g(uu^t) = v_i g$ and $v_i XX^t = v_i g g^t$. We conclude that:

$$(*) \qquad \text{For } 1 \le i \le k-2,\ v_i X^\epsilon Y^t = v_i g^\epsilon g^t.$$

Note that a_k^ϵ is unipotent, lower triangular and a_k^t is upper triangular unipotent with $(a_k^t)_{i,j} = 0$, for $j > i+1$, and $(a_k^t)_{i,i+1} = 1$. This easily implies (5), for $S = X^\epsilon Y^t$.

To show (6), note that X is given in 7.5.1, so we have $v_{k-1}X^t = v_{k-1}+v_k+\cdots+v_n$. Next, $v_{k-1}XX^t = v_{k-1}guu^tg^t = (v_{k-2}+v_{k-1})uu^tg^t = (v_{k-2}+v_{k-1}+v_{k+1})g^t = v_{k-2}+v_k+v_{k+1}+\cdots+v_n$. Also $v_{k-1}X^{-1}X^t = v_{k-1}uu^tg^{-1}g^t = (v_{k-1}+v_{k+1})g^{-1}g^t = (v_1+\cdots+v_{k-1}+v_{k+1})g^t = v_1+v_k+v_{k+1}+\cdots+v_n$.

For (7) we compute $v_{k-1}(X^t)^{-1} = v_{k-1}(g^t)^{-1}(u^t)^{-1} = (v_{k-1}+v_k)u^t = v_{k-1}+v_k+v_{k+1}+v_{k+2}$. $v_k(X^t)^{-1} = v_k(g^t)^{-1}(u^t)^{-1} = v_ku^t = v_k+v_{k+2}$ and $v_{k+1}(X^t)^{-1} = v_{k+1}(g^t)^{-1}(u^t)^{-1} = (v_{k+1}+v_{k+2})u^t = v_{k+1}+v_{k+2}$. This shows (7i). For (7ii) and (7iii), we use (7i). We compute (using (7i)) that, for $\epsilon \in \{1,-1\}$, $v_{k-1}(X^\epsilon X^t)^{-1} = (v_{k-1}+v_k+v_{k+1}+v_{k+2})X^{-\epsilon}$. If $\epsilon = 1$, we get $(v_{k-1}+v_k+v_{k+1}+v_{k+2})u^{-1}g^{-1} = (v_{k+1}+v_{k+2})g^{-1} = v_{k+2}$. If $\epsilon = -1$, we get, $(v_{k-1}+v_k+v_{k+1}+v_{k+2})gu = (v_{k-2}+v_k+v_{k+2})u = v_{k-2}+v_{k+2}$.

Next, $v_k(X^\epsilon X^t)^{-1} = (v_k+v_{k+2})X^{-\epsilon}$. If $\epsilon = 1$, we get, $(v_k+v_{k+2})u^{-1}g^{-1} = v_{k+2}g^{-1} = v_{k+1}+v_{k+2}$. If $\epsilon = -1$, we get $(v_k+v_{k+2})gu = (v_{k-1}+v_k+v_{k+1}+v_{k+2})u = v_{k+1}+v_{k+2}$.

Finally, $v_{k+1}(X^\epsilon X^t)^{-1} = (v_{k+1}+v_{k+2})X^{-\epsilon}$. If $\epsilon = 1$, we get $(v_{k+1}+v_{k+2})u^{-1}g^{-1} = (v_{k-1}+v_k+v_{k+1}+v_{k+2})g^{-1} = v_k+v_{k+2}$. If $\epsilon = -1$, we get $(v_{k+1}+v_{k+2})gu = v_{k+2}u = v_k+v_{k+2}$. This completes the proof of (7).

For (8), let $\mathcal{W} = \langle \mathcal{O}(v_1, X^t)\rangle$. By (5), $\mathcal{V}_{k-1} \subseteq \mathcal{W}$. Next, by (7i), $v_{k-1}(X^t)^{-1} = v_{k-1}+v_k+v_{k+1}+v_{k+2}$. Hence

(i) $$v + v_{k+2} \in \mathcal{W}.$$

Using (3) and 7.6 and computing modulo $\mathcal{V}_{k-1}$, $(v+v_{k+2})(X^t)^{-1} \equiv v+v_{k+2}+v_{k+3}$. Hence

(ii) $$v_{k+3} \in \mathcal{W}.$$

Now, for $k+3 \le i \le n-1$, $v_i(X^t)^{-1} = v_i + v_{i+1}$. Hence, by (ii)

(iii) $$\langle v_{k+3}, \dots, v_n\rangle \subseteq \mathcal{W}.$$

Let $\mathcal{W}' = \langle \mathcal{V}_{k-1}, v+v_{k+2}, v_{k+3}, \dots, v_n\rangle$. The reader may easily verify that $\langle v, v_{k-1}+v_k\rangle^\perp = \mathcal{W}'$ and that $vX^t = v$. We compute that $(v_{k-1}+v_k)X^t = (v_{k-1}+v_k)u^tg^t = (v_{k-1}+v_k+v_{k+1}+v_{k+2})g^t = (v_{k-1}+v_k)g^t + (v_{k+1}+v_{k+2})g^t = v_{k-1}+v_{k+1} = v+v_{k-1}+v_k$. Hence $\langle v, v_{k-1}+v_k\rangle$ is S-invariant, and it follows that $\mathcal{W}'$ is S-invariant. It follows that $\mathcal{W} = \mathcal{W}'$ and (8) is proved.

For (9), let $S = XX^t$ and set $\mathcal{W} = \langle \mathcal{O}(v_1, S)\rangle$. By (5), $\mathcal{V}_{k-1} \subseteq \mathcal{W}$. Next, by (7ii), $v_{k-1}S^{-1} = v_{k+2}$. Hence

(i′) $$v_{k+2} \in \mathcal{W}.$$

Next, we mention that all our calculations are done modulo $\mathcal{V}_{k-1}$ and we use (4) and 7.6.2. We have $v_{k+2}S^{-1} \equiv v + v_{k+3}$. Thus

(ii′) $$v + v_{k+3} \in \mathcal{W}.$$

Now $vS^{-1} = v_kS^{-1} + v_{k+1}S^{-1} = v$, by (7ii). Thus

(iii′) $$vS^{-1} = v.$$

Next $(v + v_{k+3})S^{-1} \equiv v + v_{k+2} + v_{k+4}$, hence, by (i′) and (ii′)

(iv′) $$v + v_{k+4} \in \mathcal{W}.$$

By (ii′) and (iv′)

(v′) $$v_{k+3} + v_{k+4} \in \mathcal{W}.$$

Now if $k = 4$, then $(v+v_{k+2}+v_{k+4})S^{-1} \equiv v+v+v_{k+3}+v_{k+3}+v_{k+4} = v_{k+4}$, so $v_8 \in \mathcal{W}$. It is easy to check now that by (v′), (iv′) and (ii′), (9i) holds. So from now until the end of the proof of (9) we assume that $k \geq 5$.

Next $(v + v_{k+2} + v_{k+4})S^{-1} \equiv v + v + v_{k+3} + v_{k+3} + v_{k+5} = v_{k+5}$. Thus

(vi′) $$v_{k+5} \in \mathcal{W}.$$

Suppose $k = 5$. By the above we get that $\mathcal{V}_4 \cup \{v_7, v + v_9, v_8 + v_9, v_{10}\} \subseteq \mathcal{W}$. Also, $v_{10}S^{-1} = v_9 + v_{10} \in \mathcal{W}$ and (9i) holds. So from now until the end of the proof of (9) we assume that $k \geq 6$.

Now $v_{k+5}S^{-1} \equiv v_{k+4}+v_{k+6} \in \mathcal{W}$, thus $v+v_{k+4}+v_{k+4}+v_{k+6} = v+v_{k+6} \in \mathcal{W}$, so by (ii′)

(vii′) $$v_{k+3} + v_{k+6} \in \mathcal{W}.$$

Now for $i \geq k + 3$, $(v_i + v_{i+3})S^{-1} \equiv (v_{i-1} + v_{i+2}) + (v_{i+1} + v_{i+4})$, since $v_{k+2} + v_{k+5} \in \mathcal{W}$, we conclude from (vii′) that:

(viii′) $$\text{For } k + 2 \leq i \leq n - 3,\ v_i + v_{i+3} \in \mathcal{W}.$$

Now $(v_{n-3} + v_n)S^{-1} \equiv (v_{n-4} + v_{n-1}) + (v_{n-2} + v_n)$, so from (viii′) we get

(ix′) $$v_{n-2} + v_n \in \mathcal{W}.$$

Note also that by (i′) and (viii′),

(x′) $$v_{k+j} \in \mathcal{W}, \text{ for all } 2 \leq j \leq k, \text{ such that } j \equiv 2 \pmod 3.$$

Thus, by (x′), if $k \equiv 2 \pmod 3$, $v_n \in \mathcal{W}$ and if $k \equiv 1 \pmod 3$, $v_{n-2} \in \mathcal{W}$. Thus, by (ix′), if $k \equiv 1$ or $2 \pmod 3$, $v_{n-2}, v_n \in \mathcal{W}$. It follows from (viii′) that:

(xi′) If $k \equiv 1$ or $2 \pmod 3$ then there exists $\nu \in \{0, 1\}$ such that
$$v_{k+j} \in \mathcal{W}, \text{ for all } 2 \leq j \leq k, \text{ such that } j \equiv \nu \pmod 3.$$

Since $v_{k+3} + v_{k+4} \in \mathcal{W}$, we get from (iv′), (x′), (xi′) and (viii′) that:

(xii′) If $k \equiv 1$ or $2 \pmod 3$, $\mathcal{W} \supseteq \langle \mathcal{V}_{k-1}, v_k + v_{k+1}, v_{k+2}, v_{k+3}, \dots, v_n \rangle$.

Notice that $v^{\perp} = \langle \mathcal{V}_{k-1}, v_k + v_{k+1}, v_{k+2}, v_{k+3}, \dots, v_n \rangle$ is S-invariant, as $vS = v$, so (9i) holds.

Suppose $k \equiv 0 \pmod 3$. We get from (ii′), (iv′) and (viii′), that

$$\text{(xiii}'\text{)} \qquad v + v_{k+j} \in \mathcal{W}, \text{ for all } 3 \le j \le k \text{ such that } j \equiv 0 \text{ or } 1 \pmod 3.$$

This, together with (x′), shows that

$$\mathcal{W}' := \Big\langle \mathcal{V}_{k-1}, v_{k+2}, v + v_{k+3j}, \\ v + v_{k+3j+1}, v_{k+3j+2}, v + v_n : 1 \le j \le \frac{1}{3}k - 1 \Big\rangle \subseteq \mathcal{W}.$$

It easy to check that $\langle v, v' \rangle^{\perp} = \mathcal{W}'$. We show that $v'S = v + v'$; this implies that $\langle v, v' \rangle$ is S-invariant, and hence $\mathcal{W}'$ is S-invariant, so (9ii) holds. We compute that

$$\begin{aligned} v'S &= \{(v_1 + v_3) + (v_4 + v_6) + (v_7 + v_9) + \cdots + (v_{k-2} + v_k)\} g u u^t g^t \\ &= \{(v_1 + v_2) + (v_4 + v_5) + (v_7 + v_8) + \cdots + (v_{k-2} + v_{k-1}) + v_k\} u u^t g^t \\ &= \{(v_1 + v_2) + (v_4 + v_5) + \cdots + (v_{k-2} + v_{k-1}) + v_k + v_{k+1} + v_{k+2}\} g^t \\ &= \{(v_1 + v_2) + (v_4 + v_5) + \cdots + (v_{k-2} + v_{k-1}) + v_k\} g^t \\ &\quad + (v_{k+1} + v_{k+2}) g^t \\ &= \{(v_1 + v_3) + (v_4 + v_6) + (v_7 + v_9) + \cdots + (v_{k-5} + v_{k-3}) + v_{k-2}\} \\ &\quad + v_{k+1} \\ &= v + v'. \end{aligned}$$

We now turn to the proof of (10). Set $S = X^{-1} X^t$ and $\mathcal{W} = \langle \mathcal{O}(v_1, S) \rangle$. By (5), $\mathcal{V}_{k-1} \subseteq \mathcal{W}$. Next, by (7iii), $v_{k-1} S^{-1} = v_{k-2} + v_{k+2}$. Thus

$$\text{(i}''\text{)} \qquad v_{k+2} \in \mathcal{W}.$$

Next, for $k + 2 \le i \le n - 1$, $v_i S^{-1} \equiv v_{i+1}$. Hence, by (i″)

$$\text{(ii}''\text{)} \qquad v_i \in \mathcal{W}, \quad \text{for all } \ k + 2 \le i \le n.$$

Also $v_n S^{-1} \equiv v + v_{k+2} + \cdots + v_n$, so by (ii″)

$$\text{(iii}''\text{)} \qquad v \in \mathcal{W}.$$

Again, since $v^{\perp} = \langle \mathcal{V}_{k-1}, v, v_{k+2}, v_{k+3}, \dots, v_n \rangle$ and $vS = v$, (10) holds.

7.8. *Let $1 \ne h \in C_L(X)$. Write $H = [h]_{\mathcal{C}}$ and set $Z := [X]_{\mathcal{C}}$. Write $Z = \operatorname{diag}(Z_1, Z_2)$, with $Z_1 = M_{(n-1,n,)(n-1,n)}([X]_{\mathcal{C}})$, and $Z_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. Then:*

(1) *h fixes $\langle w_1 \rangle, \langle w_1, w_2 \rangle, \dots, \langle w_1, \dots, w_{n-4} \rangle$.*
(2) *h fixes $\langle w_1, w_{n-1} \rangle$ and $\langle w_1, w_2, w_{n-1}, w_n \rangle$.*

(3) *H has the form*

$$H = \begin{bmatrix} R & E \\ F & P \end{bmatrix}$$

such that:

(3i) $I_{n-2} \neq R$ *is an* $(n-2) \times (n-2)$ *matrix commuting with* $Z, P = \begin{bmatrix} 1 & 0 \\ \delta & 1 \end{bmatrix}$, *with* $\delta \in \{0,1\}$, E *is an* $(n-2) \times 2$ *matrix whose first* $n-4$ *rows are zero, and* $E_{n-3,2} = 0$. F *is a* $2 \times (n-2)$ *matrix whose last* $n-4$ *columns are zero and* $F_{1,2} = 0$.

(3ii) $H_{i,i} = 1$, *for all* $1 \leq i \leq n$.

(3iii) *We fix the notation* $\alpha := E_{n-3,1}$, $\beta := E_{n-2,1}$, $\gamma := F_{2,1}$. *We have* $\alpha = E_{n-2,2} = F_{1,1} = F_{2,2}$.

(4) *There exists* $1 \leq r \leq n-3$, *such that* $R - I_{n-2} \in \mathcal{T}_{n-2}(r)$. *We fix the letter* r *to denote this integer.*

(5)

$$(H - I_n)^2 = \begin{bmatrix} (R - I_{n-2})^2 + EF & E' \\ F' & 0_{2,2} \end{bmatrix}$$

such that E'*is a* $(n-2) \times 2$ *matrix with* $E'_{n-2,1} = \alpha(R_{n-2,n-3} + \delta)$ (δ *as in* (3i) *and* α *as in* (3iii)) *and* $E'_{ij} = 0$ *otherwise,* F' *is a* $2 \times (n-2)$ *matrix such that* $F'_{2,1} = \alpha(R_{2,1} + \delta)$ *and* $F'_{ij} = 0$ *otherwise.* EF *is an* $(n-2) \times (n-2)$ *matrix such that* $(EF)_{n-3,1} = \alpha^2 = (EF)_{n-2,2}$, $(EF)_{n-2,1} = \alpha(\beta + \gamma)$ *and* $(EF)_{i,j} = 0$, *otherwise.*

Proof. First we mention that we think of h and H as the same linear operator, but they are distinct as matrices. The same remark holds for X and $[X]_{\mathcal{C}}$. It is easy to check that $\ker([X]_{\mathcal{C}} - I_n) = \langle w_1, w_{n-1} \rangle$, $\ker([X]_{\mathcal{C}} - I_n)^2 = \langle w_1, w_2, w_{n-1}, w_n \rangle$. Further, for $j \geq 2$, $\operatorname{im}([X]_{\mathcal{C}} - I_n)^j = \langle w_1, \ldots, w_{n-j-2} \rangle$. Thus (1) and (2) clearly hold.

Next, by (1), the first $n-4$ rows of E are zero and by (2), the last $n-4$ columns of F are zero. Also, since $\langle w_1, w_{n-1} \rangle$ is h-invariant, $F_{1,2} = 0$. Next

$$ZH = \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \cdot \begin{bmatrix} R & E \\ F & P \end{bmatrix} = \begin{bmatrix} Z_1 R & Z_1 E \\ Z_2 F & Z_2 P \end{bmatrix}$$

$$HZ = \begin{bmatrix} R & E \\ F & P \end{bmatrix} \cdot \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} = \begin{bmatrix} R Z_1 & E Z_2 \\ F Z_1 & P Z_2 \end{bmatrix}$$

so since $ZH = HZ$, R commutes with Z_1 and P commutes with Z_2. Thus $P = \begin{bmatrix} \rho & 0 \\ \mu & \rho \end{bmatrix}$. Now $(v_k + v_{k+1})h = F_{1,1} v_1 + \rho(v_k + v_{k+1})$. But $1 = Q(v_k + v_{k+1}) = Q((v_k + v_{k+1})h) = \rho^2$, so $\rho = 1$. Further, $(v_k + v_{k+1})h = F_{2,1} v_1 + F_{2,2} v_2 + \mu(v_k + v_{k+1}) + (v_k + v_{k+2})$. Hence $Q((v_k + v_{k+2})h) = \mu^2 + \nu_\epsilon + \mu$. It follows that $\nu_\epsilon = Q(v_k + v_{k+2}) = Q((v_k + v_{k+2})h) = \mu^2 + \nu_\epsilon + \mu$. Thus $\mu = 0$ or 1 and $P = \begin{bmatrix} 1 & 0 \\ \delta & 1 \end{bmatrix}$, with $\delta \in \{0,1\}$.

Next since R commutes with Z_1, 1.13 implies that, $H_{i,i} = R_{i,i} = R_{j,j} = H_{j,j}$, for all $1 \leq i, j \leq n-2$. Now

$$1 = f(v_1, v_n) = f(v_1 H, v_n H) = f(H_{1,1}v_1, H_{n-2,n-2}v_n) = H_{1,1}H_{n-2,n-2}.$$

Since $H_{1,1} = H_{n-2,n-2}$, we see that $H_{1,1} = 1$. Since $H_{n-1,n-1} = P_{1,1} = 1$ and $H_{n,n} = P_{2,2} = 1$, we see that $H_{i,i} = 1$, for all $1 \leq i \leq n$. Now since R commutes with Z_1, 1.13 implies that $R - I_{n-2} \in \mathcal{T}_{n-2}(r)$, for some $1 \leq r \leq n-3$.

Let $\begin{bmatrix} \alpha & \rho \\ \beta & \mu \end{bmatrix}$ be the last two rows of E. Then the last two rows of $Z_1 E$ are $\begin{bmatrix} \alpha & \rho \\ \alpha+\beta & \rho+\mu \end{bmatrix}$ and the last two rows of EZ_2 are $\begin{bmatrix} \alpha+\rho & \rho \\ \beta+\mu & \mu \end{bmatrix}$. Since $Z_1 E = EZ_2$, $\rho = 0$ and $\alpha = \mu$. Thus:

$$\text{The last two rows of } E \text{ are } \begin{bmatrix} \alpha & 0 \\ \beta & \alpha \end{bmatrix}.$$

Next let $\begin{bmatrix} \rho & 0 \\ \gamma & \mu \end{bmatrix}$ be the first two columns of F. Then the first two columns of $Z_2 F$ are $\begin{bmatrix} \rho & 0 \\ \rho+\gamma & \mu \end{bmatrix}$ and the first two columns of FZ_1 are $\begin{bmatrix} \rho & 0 \\ \gamma+\mu & \mu \end{bmatrix}$. Thus $\rho = \mu$. Hence:

$$\text{The first two columns of } F \text{ are } \begin{bmatrix} \rho & 0 \\ \gamma & \rho \end{bmatrix}.$$

Next $(v_k + v_{k+1})H = \rho v_1 + v_k + v_{k+1}$ and observe that $v_n H = w + v_n + \alpha(v_k + v_{k+2})$, with $w \in \langle v_1, \dots, v_{k-1}, v_k + v_{k+1}, v_{k+2}, \dots, v_{n-1} \rangle \subseteq \langle v_1, v_k + v_{k+1} \rangle^{\perp}$. Thus $0 = f(v_k + v_{k+1}, v_n) = f((v_k + v_{k+1})h, v_n h) = f(\rho v_1 + (v_k + v_{k+1}), w + v_n + \alpha(v_k + v_{k+2})) = f(\rho v_1 + (v_k + v_{k+1}), v_n + \alpha(v_k + v_{k+2})) = \rho + \alpha$. Hence $\rho = \alpha$. This completes the proof of (3) and (4), except that we must show that $R \neq I_{n-2}$. Now if $R = I_{n-2}$, then, it follows that $0 = Q(v_{n-1}) = Q(v_{n-1}H) = Q(v_{n-1} + \alpha(v_k + v_{k+1})) = \alpha$. Also, since $0 = Q(v_n) = Q(v_n H)$, $\beta = 0$. Now δ (of (3i)) must be 0; so since $h \in L$, 7.1 implies that $h = I_n$, contradicting $h \neq I_n$.

To prove (5) note that

$$(H - I_n)^2 = \begin{bmatrix} R - I_{n-2} & E \\ F & P - I_2 \end{bmatrix} \cdot \begin{bmatrix} R - I_{n-2} & E \\ F & P - I_2 \end{bmatrix} = \begin{bmatrix} (R - I_{n-2})^2 + EF & (R - I_{n-2})E + E(P - I_2) \\ F(R - I_{n-2}) + (P - I_2)F & FE + (P - I_2)^2 \end{bmatrix}.$$

Now, since the last column of $(R - I_{n-2})$ is zero, $(R - I_{n-2})E$ is an $(n-2) \times 2$ matrix, whose $(n-2, 1)$-entry is $\alpha R_{n-2,n-3}$, and whose other entries are zero. Hence it is easy to check that $E' = (R - I_{n-2})E + E(P - I_2)$, is as claimed.

Next, since the first row of $(R - I_{n-2})$ is zero, $F(R - I_{n-2})$ is a $2 \times (n-2)$ matrix whose $(2,1)$-entry is $\alpha R_{2,1}$ and whose other entries are zero. Hence, it is easy to check that $F' = F(R - I_{n-2}) + (P - I_2)F$ is as claimed. Finally, $FE = 0_{2,2}$ and clearly $(P - I_2)^2 = 0_{2,2}$. It is easy to check that EF has the claimed shape and (5) is proved.

Before formulating the next lemma it is important that the reader will recall that for a linear operator a on our vector space V, $a_{i,j}$ is the (i,j)-entry of the matrix of a, *with respect to the basis* $\mathcal{B}$, unless otherwise specified (see the beginning of Chapter 1).

7.9. *Let* $1 \neq h \in C_L(X)$. *Set* $\mathbb{T} = h - I_n$. *Write* $H = [h]_{\mathcal{C}}$. *Let* $R, P, E, F, \delta, \alpha, \beta, \gamma$ *be as in 7.8.3 and* r *as in 7.8.4. Then:*

(1) *Suppose* $k - 1 \leq r \leq n - 3$. *Then, there exists* $i \in \{1, 2\}$ *such that for* $T := \mathbb{T}^i$, *we have:*
 (1a) $\mathcal{V}_{k-1} \subseteq \ker(T)$.
 (1b) *There exists* $1 \leq f \leq n$, *such that* $T_{s,f} = 0$, *for all* $1 \leq s \leq n-1$, *and* $T_{n,f} \neq 0$.
 Further, one of the following holds.
 (1c) $\alpha \neq \delta \neq 0$, $i = 2$, $f = k+1$ *and* $\operatorname{im} T = \langle v_1, \alpha v_2 + v_k + v_{k+1} \rangle$.
 (1d) $\alpha \neq 0 = \delta$, $i = 2$, $f = 2$ *and* $\operatorname{im} T = \langle v_1, v_2 \rangle$.
 (1e) $\alpha = 0 = \delta$, $i = 1$, $f = n - r - 2$ *and*

$$\operatorname{im} T = \langle v_1, v_2, \dots, v_{n-r-2}, v_k + v_{k+1} \rangle.$$

 (1f) $\alpha = 0 = \delta$, $i = 1$, $f = n - r - 2$ *and*

$$\operatorname{im}(T) = \langle v_1, v_2, \dots, v_{n-r-3}, v_{n-r-2} + \mu(v_k + v_{k+1}) \rangle, \quad \mu \in \mathbb{F}^*.$$

 (1g) $\alpha = 0 = \delta$, $i = 1$, $f = n - r - 2$ *and* $\operatorname{im} T = \mathcal{V}_{n-r-2}$.
 (1h) $\alpha = 0 = \delta$, $r = n - 3$, $i = 1$, $f = k + 1$ *and* $\operatorname{im} T = \langle v_1, v_k + v_{k+1} \rangle$.

(2) *Suppose* $r = k - 2$ $\alpha \neq 0 = \delta$. *Then either* $\mathbb{T}^2 \in \mathcal{T}_n(n - s)$, *for some* $s \in \{1, 2\}$, *or the following holds:*
 (2a) $\mathbb{T}^2 = 0$, $H_{k-1,1} = \alpha = H_{n-2,k-1}$, $\mathcal{V}_{k-1} \subseteq \ker \mathbb{T}$, *and*
 (2b) *For all* $S \in \{X^t, XX^t, X^{-1}X^t\}$, $\ker(S\mathbb{T}) \cap \ker \mathbb{T} = \mathcal{V}_{k-2}$.

(3) *Suppose* $1 \leq r < k - 1$, *but exclude the case of (2). Then one of the following holds:*
 (3a) $r = 1$, *and* $\mathbb{T}^{n-3} \in \mathcal{T}_n(n-1)$.
 (3b) $r > 1$, $\alpha \neq 0 \neq \delta$, *and* $\ker \mathbb{T} = \{v_1, \dots, v_r, \rho v_{r+1} + \mu(v_k + v_{k+1})\}$, *with* $\rho, \mu \in \mathbb{F}^*$.
 (3c) $r = k - 2$, $\alpha \neq 0 \neq \delta$, $H_{k-1,1} = \alpha$, *and* $\ker \mathbb{T} = \mathcal{V}_{k-1}$. *Further,* $\mathbb{T}_{s,k-1} = 0$, *for all* $1 \leq s \leq n-1$, *and* $\mathbb{T}_{n,k-1} \neq 0$.
 (3d) $r = k - 2$, $\alpha \neq 0 \neq \delta$, $H_{k-1,1} = \alpha$, *and* $\operatorname{im} \mathbb{T}^2 = \langle v_1, v_k + v_{k+1} \rangle$.
 (3e) *There exists* $i \geq 1$ *and* $1 \leq m \leq k - 2$, *such that* $\operatorname{im} \mathbb{T}^i = \langle v_1, \dots, v_m \rangle$, $\mathcal{V}_{k-1} \subseteq \ker \mathbb{T}^i$ *and* $\mathbb{T}^i \in \mathcal{T}_n(n - m)$.

(3f) *There exists* $i \geq 1$*, such that* $\operatorname{im} \mathbb{T}^i = \langle v_1, \dots, v_{k-2}, v_{k-1}+v_k+v_{k+1} \rangle$ *and* $\mathcal{V}_{k-1} \subseteq \ker \mathbb{T}^i$*. Further,* $(\mathbb{T}^i)_{s,k-1} = 0$*, for all* $1 \leq s \leq n-1$*, and* $(\mathbb{T}^i)_{n,k-1} \neq 0$.

Proof. Assume the hypothesis of (1). Note that since $r \geq k-1$, $R_{2,1} = R_{n-2,n-3} = 0$. Notice also that $(R - I_{n-2})^2 = 0_{n-2,n-2}$. Suppose $\alpha \neq 0 \neq \delta$, then it is easy to verify, using 7.8.5, that (1c) holds. Similarly if $\alpha \neq 0 = \delta$, then by 7.8.5, $E' = 0_{n-2,2}$ (E' as in 7.8.5) and it is easy to verify using 7.8.5 that (1d) holds (both in the case when $\gamma = 0$ and in the case $\gamma \neq 0$). Hence we may assume that $\alpha = 0$.

We claim that:

(i) $$\text{If } r = n-3 \text{ then } \delta = 0.$$

For suppose $r = n-3$. Then $v_n H = R_{n-2,1} v_1 + v_n + \beta(v_k + v_{k+1})$. Hence $0 = Q(v_n) = Q(v_n H) = R_{n-2,1} + \beta^2$. Since by 7.8.3i, $R \neq I_{n-2}$, we get that $0 \neq R_{n-2,1} = \beta^2$. Also, $0 = f(v_n, v_k + v_{k+2}) = f(v_n H, (v_k + v_{k+2})H) = f(R_{n-2,1}v_1 + v_n + \beta(v_k + v_{k+1}), \gamma v_1 + \delta(v_k + v_{k+1}) + (v_k + v_{k+2})) = \gamma + \beta$. Hence $\gamma = \beta$. Now if $\delta = 1$, then we get that $\operatorname{im}(H - I_n) = \beta v_1 + (v_k + v_{k+1})$. But then $\dim C_V(h) = n-1$ is odd, this contradicts 7.1, since $h \in L$. So (i) holds. Further, if $r = n-3$, then, $v_n \mathbb{T} = \beta^2 v_1 + \beta(v_k + v_{k+1})$, $v_k \mathbb{T} = v_{k+1}\mathbb{T} = \beta v_1$ and $\ker \mathbb{T} = \langle \mathcal{V}_{k-1}, v_k + v_{k+1}, v_{k+2}, \dots, v_{n-1} \rangle$. Hence (1h) holds. So from now on we also assume that $k-1 \leq r < n-3$.

Note that since $\alpha = 0$, $v_k + v_{k+1} \in \ker(H - I_n)$. Hence

(ii) $$v_k \mathbb{T} = v_{k+1} \mathbb{T}$$

also, $v_k = v_{k+2} + (v_k + v_{k+2})$, so $v_k(H - I_n) = v_{k+2}(H - I_n) + (v_k + v_{k+2})(H - I_n) = H_{k,1} v_1 + \gamma v_1 + \delta(v_k + v_{k+1})$. It follows from (ii) that since $k-1 \leq r < n-3$,

(iii) $$\mathbb{T}_{k,n-r-2} = \mathbb{T}_{k+1,n-r-2} = 0.$$

Since $v_{k-1} + v_k + v_{k+1}$, $v_k + v_{k+1} \in \operatorname{Ker}(H - I_n)$, $v_{k-1} \in \operatorname{Ker} \mathbb{T}$, so since $\mathcal{V}_{k-2} \subseteq \operatorname{Ker} \mathbb{T}$, we get that $\mathcal{V}_{k-1} \subseteq \ker \mathbb{T}$, so (1a) holds. Also, since $R - I_{n-2} \in \mathcal{T}_{n-2}(r)$, and $\alpha = 0$, $v_i(H - I_n) = v_i(h - I_n) \in \mathcal{V}_{n-r-3}$, for $k+2 \leq i \leq n-1$. Thus $(h - I_n)_{i,n-r-2} = 0$, for $k+2 \leq i \leq n-1$. Finally, since $R - I_{n-2} \in \mathcal{T}_{n-2}(r)$, $H_{n-2,n-r-2} \neq 0$, so $(h - I_n)_{n,n-r-2} \neq 0$. We showed that:

(iv) If $\alpha = 0$, then $\mathbb{T}_{s,n-r-2} = 0$, for all $1 \leq s \leq n-1$, and $\mathbb{T}_{n,n-r-2} \neq 0$.

So (1b) holds for $f = n-r-2$.

Suppose $\delta \neq 0$. We leave it for the reader to verify that $\operatorname{im} \mathbb{T} = \langle v_1, v_2, \dots, v_{n-r-2}, v_k + v_{k+1} \rangle$. Hence (1e) holds.

Suppose next that $\delta = 0 = \beta$, then either $r > k-1$, in which case $\operatorname{im} \mathbb{T} = \mathcal{V}_{n-r-2}$ and (1g) holds, or $r = k-1$, in which case (1f) holds, with $\mu = 1$.

Finally suppose $\delta = 0 \neq \beta$. If $r > k-1$, then (1f) holds, with $\mu = \beta/H_{n-2,n-r-2}$, and if $r = k-1$, then either (1g) holds (in case $H_{n-2,k-1} = \beta$), or (1f) holds (otherwise). This completes the proof of (1).

Assume the hypothesis of (2). Suppose first that $(H - I_n)^2 \neq 0$. Notice that since $\delta = 0$, 7.8.5 implies that

$$(H - I_n)^2 = \begin{bmatrix} (R - I_n)^2 + EF & 0_{n-2,2} \\ 0_{2,n-2} & 0_{2,2} \end{bmatrix}.$$

Also, since $r = k-2$, $(R - I_n)^2 \in \mathcal{T}_{n-2}(n-4)$. Notice further, that by 1.13.3, $R_{r+i,i} = R_{r+s,s}$, for all $1 \leq i, s \leq n-r-2$. Thus the $(n-3,1)$-entry and the $(n-2,2)$-entry of $(R - I_n)^2$ are both equal to $R^2_{r+1,1}$. Since $(EF)_{n-3,1} = (EF)_{n-2,2} = \alpha^2$, it is clear that $(h - I_n)^2 \in \mathcal{T}_n(n-s)$, for some $s \in \{1, 2\}$.

Suppose next that $(H - I_n)^2 = 0$. Then, the above considerations imply that $R_{r+i,i} = \alpha$, for all $1 \leq i \leq n-r-2$. Note that $\mathcal{V}_{k-2} \subseteq \ker \mathbb{T}$. Also $v_{k-1}(H - I_n) = (v_{k-1} + v_k + v_{k+1})(H - I_n) + (v_k + v_{k+1})(H - I_n) = \alpha v_1 + \alpha v_1 = 0$. So $v_{k-1} \in \ker \mathbb{T}$. Thus (2a) is proved.

Next note that $\dim(\operatorname{im}(H - I_n)) = \dim(\ker(H - I_n))$, so since $(H - I_n)^2 = 0$, $\operatorname{im}(H - I_n) = \ker(H - I_n)$. Also $v_n(H - I_n) = v' + R_{n-2,k-1}(v_{k-1} + v_k + v_{k+1}) + \alpha v_{k+2} + \beta(v_k + v_{k+1}) + \alpha(v_k + v_{k+2}) = v'' + \alpha v_k + (R_{n-2,k-1} + \beta)(v_k + v_{k+1})$, with $v' \in \mathcal{V}_{k-2}$ and $v'' \in \mathcal{V}_{k-1}$. Hence

(v) $\quad v_n(H - I_n) \equiv \alpha v_k + (R_{n-2,k-1} + \beta)(v_k + v_{k+1}) (\operatorname{mod} \mathcal{V}_{k-1}).$

Since $\mathcal{V}_{k-1} \subseteq \ker(H - I_n)$, we get from (v) that

(vi) $\quad \rho v_k + \mu v_{k+1} \in \ker(H - I_n), \quad$ for some $\mu, \rho \in \mathbb{F}$, with $\mu \neq \rho$.

Thus

(vii) $\quad \ker \mathbb{T} = \langle \mathcal{V}_{k-1}, \rho v_k + \mu v_{k+1} \rangle \quad \rho, \mu$ as in (vi).

For (2b), we'll show that if ρ, μ are as in (vi) and $S \in \{X^t, XX^t, X^{-1}X^t\}$, $\langle v_{k-1}S^{-1}, (\rho v_k + \mu v_{k+1})S^{-1} \rangle \cap \ker \mathbb{T} = (0)$. This easily implies $\ker \mathbb{T} \cap \ker S\mathbb{T}$ has dimension $\leq k-2$. Since, by (vii) and 7.7.5, $\mathcal{V}_{k-2} \subseteq \ker \mathbb{T} \cap \ker S\mathbb{T}$, (2b) follows. Let $v \in \langle v_{k-1}S^{-1}, (\rho v_k + \mu v_{k+1})S^{-1} \rangle$.

Suppose $S = X^t$. By 7.7.7i, $v = \theta_1 v_{k-1}S^{-1} + \theta_2(\rho v_k + \mu v_{k+1})S^{-1} = \theta_1(v_{k-1} + v_k + v_{k+1} + v_{k+2}) + \theta_2(\rho(v_k + v_{k+2}) + \mu(v_{k+1} + v_{k+2})) = \theta_1 v_{k-1} + (\theta_1 + \theta_2\rho)v_k + (\theta_1 + \theta_2\mu)v_{k+1} + (\theta_1 + \theta_2(\rho + \mu))v_{k+2}$. So if $v \in \ker \mathbb{T}$, then, by (vii), $\theta_1 + \theta_2(\rho + \mu) = 0$. Thus, $\theta_1 + \theta_2\rho = \theta_2\mu$ and $\theta_1 + \theta_2\mu = \theta_2\rho$. It follows that $\theta_2\mu v_k + \theta_2\rho v_{k+1} \in \ker \mathbb{T}$. Hence, we may assume that $\theta_2\mu v_k + \theta_2\rho v_{k+1} = \rho v_k + \mu v_{k+1}$. Hence $\theta_2\mu + \rho = \theta_2\rho + \mu = 0$. This is possible only if $\rho = \mu$, a contradiction.

Suppose $S = XX^t$. Then, by 7.7.7ii, $v = \theta_1 v_{k-1}S^{-1} + \theta_2(\rho v_k + \mu v_{k+1})S^{-1} = \theta_1 v_{k+2} + \theta_2\{\rho(v_{k+1} + v_{k+2}) + \mu(v_k + v_{k+2})\} = \theta_2\mu v_k + \theta_2\rho v_{k+1} + (\theta_1 + \theta_2(\rho + \mu))v_{k+2}$. So if $v \in \ker(h - I_n)$, then, by (vii), $\theta_1 + \theta_2(\rho + \mu) = 0$ and $\theta_2\mu v_k + \theta_2\rho v_{k+1} \in \ker \mathbb{T}$, which we have seen to be impossible.

Suppose $S = X^{-1}X^t$. Then, by 7.7.7iii, $v = \theta_1 v_{k-1}S^{-1} + \theta_2(\rho v_k + \mu v_{k+1})S^{-1} = \theta_1(v_{k-2} + v_{k+2}) + \theta_2(\rho(v_{k+1} + v_{k+2}) + \mu(v_k + v_{k+2}))$ and as in the case $S = XX^t$, we get a contradiction. This completes the proof of (2).

Assume the hypothesis of (3).

Case 1. $r = 1$.

By 7.8.5, $(H - I_n)^2 = \begin{bmatrix} t & E' \\ F' & 0_{2,2} \end{bmatrix}$, with $t \in \mathcal{T}_{n-2}(2)$. Then, it is easy to verify that $(H - I_n)^3 = \begin{bmatrix} t' & 0 \\ 0 & 0_{2,2} \end{bmatrix}$, with $t' \in \mathcal{T}_{n-2}(3)$ and from that (3a) follows easily.

So from now on we assume that $r > 1$.

Case 2. $\alpha \neq 0 \neq \delta$.

If $r \neq k-2$, or $r = k-2$ and $H_{k-1,1} \neq \alpha$, then it is easily checked that (3b) holds. So suppose that $r = k-2$, and $H_{k-1,1} = \alpha$. Then $v_{k-1}\mathbb{T} = (v_{k-1} + v_k + v_{k+1})\mathbb{T} + (v_k + v_{k+1})\mathbb{T} = \alpha v_1 + \alpha v_1 = 0$. So clearly $\ker \mathbb{T} = \mathcal{V}_{k-1}$. Also, for $k+2 \leq s \leq n-2$, $v_s\mathbb{T} \in \mathcal{V}_{k-2}$. Further, $(v_k + v_{k+2})\mathbb{T} = \gamma v_1 + \alpha v_2 + (v_k + v_{k+1})$ and $v_{k+2}\mathbb{T} = R_{k,1}v_1 + R_{k,2}v_2$. Since $v_k\mathbb{T} = v_{k+2}\mathbb{T} + (v_k + v_{k+2})\mathbb{T}$, we conclude that $\mathbb{T}_{k,k-1} = 0$. Also since $(v_k + v_{k+1})\mathbb{T} = \alpha v_1$, we see that $\mathbb{T}_{k+1,k-1} = 0$. Hence, we see that $\mathbb{T}_{s,k-1} = 0$, for all $1 \leq s \leq n-1$. Now $v_n\mathbb{T} = v' + R_{n-2,k-1}(v_{k-1} + v_k + v_{k+1}) + R_{n-2,k}v_{k+2} + \beta(v_k + v_{k+1}) + \alpha(v_k + v_{k+1})$, with $v' \in \mathcal{V}_{k-2}$. Hence, if $R_{n-2,k-1} \neq 0$, then $\mathbb{T}_{n,k-1} \neq 0$, and case (3c) holds. Finally, suppose $R_{n-2,k-1} = 0$. Then $v_nh = v_nH = v'' + v_n + \beta w_{n-1} + \alpha w_n$, with $v'' \in \langle \mathcal{V}_{k-2}, v_{k+2} \rangle$ and $w_nh = \gamma v_1 + \alpha v_2 + w_{n-1} + w_n$. Hence $0 = f(v_n, w_n) = f(v_nh, w_nh) = \gamma + \beta + \alpha$. Hence $\beta + \gamma = \alpha$. Also, $v_{k+2}h = R_{k,1}v_1 + R_{k,2}v_2 + v_{k+2}$. Hence, $0 = f(v_{k+2}, v_n) = f(v_{k+2}h, v_nh) = R_{k,1}$. So $R_{k,1} = 0$. Since $\beta + \gamma = \alpha$, 7.8.5 yields $(EF)_{n-2,1} = \alpha^2$. Then, since $R_{k,1} = R_{n-2,k-1} = 0$ and $R_{k-1,1} = R_{n-2,k} = \alpha$ (see 1.13.3), we get, using 7.8.5, that $(R - I_{n-2})^2 + EF \in \mathcal{T}_{n-2}(n-3)$. Now using 7.8.5, it is easy to check that (3d) holds.

Case 3. $\alpha \neq 0 = \delta$ and $r \neq k-2$; or $\alpha = 0$.

Using 7.8.5 we get that

$$(H - I_n)^2 = \begin{bmatrix} (R - I_n)^2 + EF & 0_{n-2,2} \\ 0_{2,n-2} & 0_{2,2} \end{bmatrix}.$$

Now if $\alpha = 0$, $EF = 0$, while if $\alpha \neq 0 = \delta$, and $r \neq k-2$, then $(R - I_n)^2 + EF \in \mathcal{T}_{n-2}(r')$, for some $1 \leq r' < n-2$. Thus in either case

$$(H - I_n)^2 = \begin{bmatrix} t & 0_{n-2,2} \\ 0_{2,n-2} & 0_{2,2} \end{bmatrix}$$

with $t \in \mathcal{T}_{n-2}(r')$, for some $1 \leq r' < n-2$. It follows that for some i,

$$(H - I_n)^i = \begin{bmatrix} t' & 0_{n-2,2} \\ 0_{2,n-2} & 0_{2,2} \end{bmatrix}$$

with $t' \in \mathcal{T}_{n-2}(r'')$, for some $k-1 \leq r'' < n-2$. If $r'' > k-1$, we get case (3e). So suppose $r'' = k-1$. Clearly, $\operatorname{im}\mathbb{T}^i = \langle v_1, \dots, v_{k-2}, v_{k-1}+v_k+v_{k+1}\rangle$ and $\mathcal{V}_{k-1} \subseteq \ker\mathbb{T}^i$. So, to establish (3f), it remains to show that $(\mathbb{T}^i)_{s,k-1} = 0$, for all $1 \leq s \leq n-1$, and $(\mathbb{T}^i)_{n,k-1} \neq 0$. Now for $k+2 \leq s \leq n-1$, $v_s(H-I_n)^i \in \mathcal{V}_{k-2}$, so $(\mathbb{T}^i)_{s,k-1} = 0$. Further since $(v_k + v_{k+1})\mathbb{T}^i = (v_k + v_{k+2})\mathbb{T}^i = 0$, $v_k\mathbb{T}^i = v_{k+1}\mathbb{T}^i = v_{k+2}\mathbb{T}^i \in \langle v_1\rangle$. Hence $(\mathbb{T}^i)_{k,k-1} = (\mathbb{T}^i)_{k+1,k-1} = 0$. Finally, since $t' \in \mathcal{T}_{n-2}(k-1)$, $(\mathbb{T}^i)_{n,k-1} \neq 0$. Thus, (3f) holds.

7.10. *Let $\epsilon \in \{-1, 1\}$ and let $S \in \{X^t, X^\epsilon X^t\}$. Let $R \in C_L(S)$ and suppose v_1 is a characteristic vector of R. Then $R = 1$.*

Proof. Set $\mathcal{W} = \langle \mathcal{O}(v_1, S)\rangle$. Using, 7.7.8, 7.7.9 and 7.7.10, it is clear that $\mathcal{W}$ is nonsingular (in all cases) and hence R centralizes $\mathcal{W}$. Set $v = v_k + v_{k+1}$.

Suppose first that $S = X^t$. Then, by 7.7.8, $\mathcal{W}^\perp = \langle v, v'\rangle$, with $v' = v_{k-1} + v_k$, $vS = v$ and $v'S = v + v'$. Clearly $\mathcal{W}^\perp$ is R-invariant and since $R \in C_L(S)$, $vR = \alpha v$ and $v'R = \beta v + \alpha v'$. Since $Q(v) = 1$, $\alpha = 1$. Hence R centralizes $\langle \mathcal{W}, v\rangle$ of dimension $n-1$. Thus, by 7.1 (and since $\det(R) = 1$), $R = 1$.

Suppose next that $S = XX^t$ and that $k \equiv 0 \pmod 3$. Then using 7.7.9 and arguing exactly as in previous paragraph we get $R = 1$.

Finally suppose $S = XX^t$ and $k \not\equiv 0 \pmod 3$, or $S = X^{-1}X^t$. By 7.7.9 and 7.7.10, $\dim(\mathcal{W}) = n-1$, so by 7.1, $R = 1$.

7.11. *Let $\epsilon \in \{1, -1\}$ and let $S \in \{X^t, X^\epsilon X^t\}$. Suppose $R \in \Delta^{\leq 2}(X) \cap \Delta^{\leq 1}(S)$. Then v_1 is a characteristic vector of R.*

Proof. Let $h \in \Delta^{\leq 1}(X) \cap \Delta^{\leq 1}(R)$. We'll show that there exists $i \geq 1$, such that if we set $T = (h - I_n)^i$, then there are integers $j, m, \ell \geq 0$ such that all the hypotheses of 1.15 are satisfied for S, T and R. The lemma will follow from 1.15. We'll use 7.9, so we adopt the notation of 7.9. For a subspace $\mathcal{W} \subseteq V$, let $\mathfrak{S}(\mathcal{W}) = \langle w \in \mathcal{W} : Q(w) = 0\rangle$ (the singular vectors of $\mathcal{W}$).

Case 1. $k-1 \leq r \leq n-3$.

In each case (1c)-(1h) of 7.9.1 we pick i as defined in these cases. We take $j = k-2$, in all cases. Notice that by 7.7.5, hypothesis (a) of 1.15 is satisfied. We let $m = \dim\{\mathfrak{S}(\operatorname{im} T)\}$ and $\ell = 1$. Using 7.7.6 and (1b) of 7.9.1, we get hypothesis (c) of 1.15. The remaining hypotheses of 1.15 are readily verified using 7.9.1.

Case 2. $r = k-2$ and $\alpha \neq 0 = \delta$.

In this case, if $(h - I_n)^2 \in \mathcal{T}_n(n-s)$, for some $s \in \{1, 2\}$, we take $i = 2$, $m = s$, $j = k-2$ and $\ell = 1$. Otherwise we take $i = 1$, $j = k-2 = m$ and $\ell = 1$. Using 7.9.2, we see that the hypotheses of 1.15 are satisfied.

Case 3. $1 \leq r < k-1$, but Case 2 does not occur.

If case 7.9.3a holds, take $i = n-3$ and $m = 1$, to get the lemma trivially. If case 7.9.3b holds, take $i = 1$, $j = m = \dim(\mathfrak{S}(\ker T))$ and $\ell = 0$. If case 7.9.3c holds, take $i = 1$, $j = k - 2$, $m = k - 1$ and $\ell = 1$. Notice again that by 7.7.6, hypothesis (c) of 1.15 holds. If case 7.9.3d holds, then $\mathfrak{S}(\mathrm{im}\,(h - I_n)^2) = \langle v_1 \rangle$ and trivially, $\langle v_1 \rangle$ is R-invariant. If case 7.9.3e holds, take i as in 7.9.3e, $j = k - 2$, m as in 7.9.3e and $\ell = 1$. If case 7.9.3f holds, take i as in 7.9.3f, $j = k - 2$, $m = \dim\{\mathfrak{S}(\mathrm{im}\,(T)\} = k - 2$, and $\ell = 1$. Using 7.7.6, the hypotheses of 1.15 are readily verified in cases 7.9.3e and 7.9.3f and the proof of 7.11 is complete.

7.12. *Let $\Lambda = \Delta(L)$, $\epsilon \in \{1, -1\}$ and let $S \in \{X^t, X^\epsilon X^t\}$. Then $d_\Lambda(X, S) \geq 4$.*

Proof. Suppose $d_\Lambda(X, S) \leq 3$ and let $R \in \Delta^{\leq 2}(X) \cap \Delta^{\leq 1}(S)$. By 7.11, v_1 is a characteristic vector of R and by 7.10, $R = 1$, a contradiction.

Theorem 7.13. *$\Delta(L)$ is balanced.*

Proof. Let $\Lambda = \Delta(L)$. Note that 7.12 implies that $B_\Lambda(X, X^t)$ and by 1.9, $B_\Lambda(X^t, X)$, so Λ is balanced.

Chapter 2. The Exceptional Groups of Lie type.

In Section 8 we prove that for all exceptional groups of Lie type L excluding $E_7(q)$, the commuting graph $\Delta(L)$ is disconnected (Theorem 8.8). In Section 9 we prove that if $L \cong E_7(q)$, then $\Delta(L)$ is balanced (see 1.3.2).

8. The Exceptional Groups excluding $E_7(q)$.

In this section L is a finite exceptional group of Lie type, excluding $E_7(q)$. We take $L = G_\sigma$, where G is a simply connected simple algebraic group and σ is a Frobenius morphism. Hence L is one of the following groups: ${}^2B_2(2^{2m+1})$, $G_2(q)$, ${}^2G_2(3^{2m+1})$, ${}^3D_4(q)$, $F_4(q)$, ${}^2F_4(2^{2m+1})$, $E_6(q)$, ${}^2E_6(q)$, $E_8(q)$. We exclude certain small cases where L is either solvable or L' is of classical type. So we exclude ${}^2B_2(2)$, $G_2(2)$,${}^2G_2(3)$. The remaining groups are all quasisimple, with the exception of ${}^2F_4(2)$, which has derived group of index 2. We let $L^* = L/Z(L)$. Of course $Z(L) = 1$, except when $L \cong E_6(q)$, in which case $|Z(L)| = (3, q - 1)$, and when $L \cong {}^2E_6(q)$, in which case $|Z(L)| = (3, q + 1)$.

8.1. *Assume G is a simply connected simple algebraic group and σ is a Frobenius morphism with quasisimple fixed point group G_σ. Let T be a σ invariant maximal torus. Suppose $s \in T_\sigma$ is an element such that $s \notin S_\sigma$, for any σ-invariant maximal torus S, such that $|S_\sigma| \neq |T_\sigma|$. Then $C_{G_\sigma}(s) = T_\sigma$.*

Proof. It will suffice to show that $C_G(s) = T$. As G is simply connected, $C_G(s) = C_G(s)^0$ ([**1**, II, 3.9]) and this is a reductive group. Write $C_G(s) = DZ$, where $Z = Z(C_G(s))^0$ and $D = C_G(s)'$. Thus D is a semisimple group. Note that $T \leq C_G(s)$ and that s is contained in all maximal tori of $C_G(s)$ (as maximal tori are self centralizing).

If $D = 1$, then $C_G(s) = T$, as required. Suppose this is not the case and let $\{D_1, \dots, D_r\}$ be an orbit of $\langle\sigma\rangle$ on simple components of D. Then σ^r induces a Frobenius morphism on each D_i. By [**1**, I, 2.9], this Frobenius morphism normalizes a maximal torus contained in an invariant Borel of D_1. Taking images under powers of σ we get a maximal torus of each D_i with the same properties.

For the moment exclude the case where $p = 2$ and $D_i = B_2, C_2$. Then σ^r acts on the various root systems, stabilizing the positive roots, and fixing the root of highest height and its negative. Hence for each i, σ^r normalizes J_i, the fundamental SL_2 generated by the corresponding root subgroups. Also σ normalizes $J_1 \cdots J_r$. The centralizer in $C_G(s)$ of this group is also σ-stable and so contains a σ-stable maximal torus, say E.

There are two classes of σ-invariant maximal tori in $J_1 \cdots J_r$. These correspond to maximal tori in the fixed point group (of type $A_1(q^r)$ of order $q^r + 1$ and $q^r - 1$). Hence there are two classes of σ-invariant maximal tori of $(J_1 \cdots J_r)E$ whose fixed points in $J_1 \cdots J_r$ have order $q^r + 1$ and $q^r - 1$. A representative of one of these tori, say $\bar{T}$ has fixed points of order different than that of T_σ, however, by earlier remarks, $s \in \bar{T}_\sigma$, contradicting the hypothesis.

Finally consider the case $p = 2$, and $D_i = B_2, C_2$. This is only possible when $G = F_4$. There cannot be more than one such simple component in D, since the product of two has trivial centralizer, so cannot lie in $C_G(s)$. Thus D_1 is σ-invariant and we can use the same argument unless $(D_1)_\sigma = Sz(q)$. Here too there are at least two classes of maximal tori, so we can proceed as above.

Corollary 8.2. *Let G be a simple connected simple algebraic group and let σ be a Frobenius morphism of G such that $G_\sigma = L$. Let T be a σ-invariant torus and assume:*

(a) *If $S \leq G$ is a σ-invariant maximal torus such that $|S_\sigma| \neq |T_\sigma|$, then $(|T_\sigma|, |S_\sigma|) = |Z(L)|$.*
(b) $(|T_\sigma : Z(L)|, |Z(L)|) = 1$.

Let T^*_σ be the image of T_σ in L^*. Then $T^*_\sigma - \{1\}$ is a component of $\Delta(L^*)$.

Proof. We'll show that $C_{L^*}(s) = T^*_\sigma$, for every $1 \neq s^* \in T^*_\sigma$. Let $s \in T_\sigma - Z(L)$. We claim that $s \notin S_\sigma$, for every σ-invariant maximal torus S of G, such that $|S_\sigma| \neq |T_\sigma|$. Indeed, since $s \in T_\sigma - Z(L)$, (b) implies that

$|s| \nmid |Z(L)|$, where $|s|$ is the order of s. However, if $s \in S_\sigma$, for some σ-invariant maximal torus S of G, then $|s|$ divides $(|T_\sigma|, |S_\sigma|)$. Hence, by (a), $|S_\sigma| = |T_\sigma|$.

By 8.1, $C_L(s) = T_\sigma$. Hence, from (b) we get that $C_{L^*}(s^*) = T^*_\sigma$.

Notation and definitions. We denote by $\Phi_n(x)$, the n-th cyclotomic polynomial (of degree $\phi(n)$). Given a prime p and an integer b, the *p-share* of b is the largest power of p dividing b.

8.3. *Let $n, a \geq 2$ and let p be a prime. When $(a,p) = 1$, denote by $d_p(a)$ the order of a mod p. Then:*

(1) *$p \mid \Phi_n(a)$ iff $(a,p) = 1$, and $n = p^e d_p(a)$, for some $e \geq 0$.*
(2) *If $n \geq 3$, and $p \mid \Phi_n(a)$, then either $n = d_p(a)$, or the p-share of $\Phi_n(a)$ is p.*

Proof. This is well-known, see, e.g., [**9**, p. 27].

Corollary 8.4. *Let r be a prime, q a positive power of r and $2 \leq m < n$. Then:*

(1) *If $m \nmid n$ or if $\frac{n}{m}$ is not a prime power, then $(\Phi_n(q), \Phi_m(q)) = 1$.*
(2) *If $\frac{n}{m} = p^f$, with $r \neq p$ a prime and $f \geq 1$, then $(\Phi_n(q), \Phi_m(q)) = p^t$, with $t \geq 0$.*

Proof. Let p be a prime such that $p \mid (\Phi_n(q), \Phi_m(q))$. By 8.3.1, $p \neq r$, $m = p^{e_1} d_p(q)$ and $n = p^{e_2} d_p(q)$. Thus $m \mid n$ and $\frac{n}{m} = p^{e_2 - e_1}$. This shows (1). It also shows (2), since, we just saw that there can be at most one prime dividing $(\Phi_n(q), \Phi_m(q))$.

In the following lemma we list the cyclotomic polynomials of degree ≤ 8. These are the relevant cyclotomic polynomials in calculating the order of maximal tori in exceptional groups of Lie type.

8.5. *The cyclotomic polynomials of degree ≤ 8 are given in the following table.*

The degree	*The cyclotomic polynomials*
1	$\Phi_1(x) = x - 1, \quad \Phi_2(x) = x + 1.$
2	$\Phi_3(x), \quad \Phi_4(x) = x^2 + 1, \quad \Phi_6(x) = x^2 - x + 1.$
4	$\Phi_5(x), \quad \Phi_8(x) = x^4 + 1, \quad \Phi_{10}(x) = x^4 - x^3 + x^2 - x + 1,$ $\Phi_{12}(x) = x^4 - x^2 + 1.$
6	$\Phi_7(x), \quad \Phi_9(x) = x^6 + x^3 + 1,$ $\Phi_{14}(x) = x^6 - x^5 + x^4 - x^3 + x^2 - x + 1,$ $\Phi_{18}(x) = x^6 - x^3 + 1.$
8	$\Phi_{15}(x) = x^8 - x^7 + x^5 - x^4 + x^3 - x + 1, \quad \Phi_{16}(x) = x^8 + 1,$ $\Phi_{20}(x) = x^8 - x^6 + x^4 - x^2 + 1, \quad \Phi_{24}(x) = x^8 - x^4 + 1,$ $\Phi_{30} = x^8 + x^7 - x^5 - x^4 - x^3 + x + 1.$

Proof. The degree of $\Phi_n(x)$ is $\phi(n) = \prod_{i=1}^{k} p_i^{m_i - 1}(p_i - 1)$, where $n = \prod_{i=1}^{k} p_i^{m_i}$ and it is easy to calculate the table.

Corollary 8.6. *Let q be a positive power of a prime r. Then:*

(1) $(\Phi_{12}(q), f(q)) = 1$, *for any cyclotomic polynomial $f(x)$ of degree ≤ 4 distinct from $\Phi_{12}(x)$.*

(2) *Let $f(x)$ be a cyclotomic polynomial of degree ≤ 6, distinct from $\Phi_9(x)$. Then:*
  - (i) *If $f(x) \notin \{\Phi_1(x), \Phi_3(x)\}$, then $(\Phi_9(q), f(q)) = 1$.*
  - (ii) *The 3-share of $\Phi_9(q)$ is $(3, q - 1)$.*
  - (iii) *If $f(x) \in \{\Phi_1(x), \Phi_3(x)\}$, then $(\Phi_9(q), f(q)) = (3, q - 1)$.*

(3) *Let $f(x)$ be a cyclotomic polynomial of degree ≤ 6, distinct from $\Phi_{18}(x)$. Then:*
  - (i) *If $f(x) \notin \{\Phi_2(x), \Phi_6(x)\}$, then $(\Phi_{18}(q), f(q)) = 1$.*
  - (ii) *The 3-share of $\Phi_{18}(q)$ is $(3, q + 1)$.*
  - (iii) *If $f(x) \in \{\Phi_2(x), \Phi_6(x)\}$, then $(\Phi_{18}(q), f(q)) = (3, q + 1)$.*

(4) $(\Phi_{30}(q), f(q)) = 1$, *for any cyclotomic polynomial $f(x)$, of degree ≤ 8, distinct from $\Phi_{30}(x)$.*

(5) *Let $f(x)$ be a cyclotomic polynomial of degree ≤ 6, distinct from $\Phi_{14}(x)$. Then:*
  - (i) *If $f(x) \neq \Phi_2(x)$, then $(\Phi_{14}(q), f(q)) = 1$.*
  - (ii) $(\Phi_{14}(q), \Phi_2(q)) = (q + 1, 7)$.

(6) *Let $f(x)$ be a cyclotomic polynomial of degree ≤ 6, distinct from $\Phi_7(x)$. Then:*
  - (i) *If $f(x) \neq x - 1$, then $(\Phi_7(q), f(q)) = 1$.*
  - (ii) $(\Phi_7(q), q - 1) = (q - 1, 7)$.

Proof. (1): We have $\Phi_{12}(x) = x^4 - x^2 + 1$, hence clearly $(\Phi_{12}(q), \Phi_1(q)) = 1$. Let $\Phi_{12}(x) \neq f(x)$ be a cyclotomic polynomial of degree ≤ 4. Note that $\Phi_{12}(q)$ is odd and $\Phi_{12}(q) \equiv 1 \pmod 3$. Now, by 8.5, $f(x) = \Phi_m(x)$, with $m < 12$, so (1) follows from 8.4.

(2): Next $\Phi_9(x) = q^6 + q^3 + 1$. Let $\Phi_9(x) \neq f(x)$ be a cyclotomic polynomial of degree ≤ 6. Since $\Phi_9(q)$ is odd, 8.4 implies that $(\Phi_9(q), \Phi_{18}(q)) = 1$. Now, by 8.5 and 8.4, $(\Phi_9(q), f(q)) = 1$, except when $q \equiv 1 \pmod 3$ and $f(x) = \Phi_1(x)$ or $\Phi_3(x)$, in which case $(\Phi_9(q), f(q)) = 3^t$, for some $t \geq 1$. Suppose $q \equiv 1 \pmod 3$, then $d_3(q) = 1$, so by 8.3.2, the 3-share of $\Phi_9(q)$ is 3 and (2) follows.

(3): Next, $\Phi_{18}(x) = x^6 - x^3 + 1$. We already observed that $(\Phi_{18}(q), \Phi_9(q)) = 1$. Let $\Phi_{18}(x) \neq f(x)$ be a cyclotomic polynomial of degree ≤ 6. Notice that $(\Phi_{18}(q), \Phi_1(q)) = 1$. Since $\Phi_{18}(q)$ is odd, 8.5 and 8.4 imply that, $(\Phi_{18}(q), f(q)) = 1$, except when $f(x) = \Phi_2(x)$ or $\Phi_6(x)$ and $q \equiv -1 \pmod 3$, in which case $(\Phi_{18}(q), f(q)) = 3^t$, for some $t \geq 1$. But by 8.3.2, if $q \equiv -1 \pmod 3$, the 3-share of $\Phi_{18}(q)$ is 3 and (3) holds.

(4): Let $\Phi_{30}(x) \neq f(x)$ be a cyclotomic polynomial of degree ≤ 8 and suppose $(\Phi_{30}(q), f(q)) \neq 1$. Now $\Phi_{30}(x) = x^8 + x^7 - x^5 - x^4 - x^3 + x + 1$, so $\Phi_{30}(q)$ is odd. Notice that $(\Phi_{30}(q), \Phi_1(q)) = 1$. By 8.5 and 8.4, $f(x) = \Phi_m(x)$ for some $1 < m < 30$. By 8.4, if p is a prime dividing $(\Phi_{30}(q), f(q))$, then $p = 3$ or 5. Now by 8.3.1, $\Phi_{30}(q) \not\equiv 0 \pmod 3$ and $\Phi_{30}(q) \not\equiv 0 \pmod 5$ so (4) follows.

(5): Let $\Phi_{14}(x) \neq f(x)$ be a cyclotomic polynomial of degree ≤ 6 and suppose $(\Phi_{14}(q), f(q)) \neq 1$. Now $\Phi_{14}(x) = x^6 - x^5 + x^4 - x^3 + x^2 - x + 1$, so $\Phi_{14}(q)$ is odd. Using 8.5 and 8.4, we see that $f(x) = \Phi_2(x)$ and $(\Phi_{14}(q), \Phi_2(q)) = 7^t$, for some $t \geq 1$. Hence $q \equiv -1 \pmod 7$ and by 8.3.2, $t = 1$.

(6): Let $\Phi_7(x) \neq f(x)$ be a cyclotomic polynomial of degree ≤ 6 and suppose $(\Phi_7(q), f(q)) \neq 1$. Now $\Phi_7(x) = x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$, so $\Phi_7(q)$ is odd. Using 8.5 and 8.4, we see that $f(x) = x - 1$. Now $\Phi_7(x) = (x^5 + 2x^4 + 3x^3 + 4x^2 + 5x + 6)(x - 1) + 7$. Hence $(\Phi_7(q), q - 1) = (q - 1, 7)$.

8.7. *There exists a maximal torus $T_\sigma \leq L$ satisfying the hypotheses of* 8.2.

Proof. We begin with the Suzuki and Ree groups ${}^2B_2(q)$, ${}^2G_2(q)$, ${}^2F_4(q)$, where $p = 2, 3, 2$ respectively. Here $q = p^{2m+1}$ and we set $q_0 = \sqrt{q}$. Suppose first that $L \simeq {}^2B_2(q)$. As is well-known, (see, e.g., [**1**, p. 191]) there are 3 classes of maximal tori in L of orders $(q-1)$, $(q-\sqrt{2q}+1)$ and $(q+\sqrt{2q}+1)$. So taking, e.g., $|T_\sigma| = q - 1$, we are done.

Suppose next that $L \cong {}^2G_2(q)$. Then, there are 4 classes of maximal tori in L (see, e.g., [**1**, p. 213]) of orders $(q - 1)$, $(q + 1)$, $q - \sqrt{3q} + 1$ and $q + \sqrt{3q} + 1$ and taking, e.g., $|T_\sigma| = q + \sqrt{3q} + 1$, we are done.

Suppose that $L \cong {}^2F_4(q)$. By [**17**], the order of a maximal torus of L either divides $[\Phi_1(q)]^2[\Phi_2(q)]^2\Phi_4(q)\Phi_6(q)$, or is of order $q_0^4+\epsilon\sqrt{2}q_0^3+q_0^2+\epsilon\sqrt{2}q_0+1$,

$\epsilon \in \{1, -1\}$ and hence divides $\Phi_{12}(q)$. Let $|T_\sigma| = q_0^4 + \sqrt{2}q_0^3 + q_0^2 + \sqrt{2}q_0 + 1$ and let $S_\sigma \leq L$ be a maximal torus with $|S_\sigma| \neq |T_\sigma|$. Since $|T_\sigma|$ divides $\Phi_{12}(q)$, we deduce from 8.6.1, that $(|T_\sigma|, |S_\sigma|) = 1$, except perhaps when $|S_\sigma| = q_0^4 - \sqrt{2}q_0^3 + q_0^2 - \sqrt{2}q_0 + 1$. But it is easy to check that $(q_0^4 + \sqrt{2}q_0^3 + q_0^2 + \sqrt{2}q_0 + 1, q_0^4 - \sqrt{2}q_0^3 + q_0^2 - \sqrt{2}q_0 + 1) = 1$.

Suppose L is one of the remaining types. Let $S_\sigma \leq L$ be a maximal torus. As is well-known, if n is the rank of L, then

$$(*) \qquad |S_\sigma| = g(q)$$

where $g(x)$ is a polynomial of degree n, a product of cyclotomic polynomials.

If $L \cong G_2(q)$, with $q \not\equiv -1 \pmod 3$ we let $|T_\sigma| = \Phi_6(q)$, while if $q \equiv -1 \pmod 3$, we let $|T_\sigma| = \Phi_3(q)$. If $L \cong {}^3D_4(q)$, we let $|T_\sigma| = \Phi_{12}(q)$. If $L \cong F_4(q)$, we let $|T_\sigma| = \Phi_{12}(q)$. If $L \cong E_6(q)$ we let $|T_\sigma| = \Phi_9(q)$. If $L \cong {}^2E_6(q)$, we let $|T_\sigma| = \Phi_{18}(q)$. Finally, if $L \cong E_8(q)$, we let $|T_\sigma| = \Phi_{30}(q)$.

In all cases T_σ exists (see, e.g., [**1**, pp. 304-305] and [**5**]). By 8.6 and $(*)$, T_σ satisfies the hypotheses of 8.2.

Theorem 8.8. *Let L^* be an exceptional finite simple group of Lie type. Suppose L^* is not of type E_7. Then $\Delta(L^*)$ is disconnected.*

Proof. This is immediate from 8.2 and 8.7.

9. The group $E_7(q)$.

In this section q is a prime power and L is a simple group with $L \cong E_7(q)$. We let $\delta = \gcd(q-1, 2)$. Recall that

$$|L| = \frac{1}{\delta} q^{63}(q^{18}-1)(q^{14}-1)(q^{12}-1)(q^{10}-1)(q^8-1)(q^6-1)(q^2-1).$$

Thus if $\tilde{L}$ is the universal group of type E_7 defined over the field of q elements, then $|Z(\tilde{L})| = \delta$ and $\tilde{L}/Z(\tilde{L}) = L$. We let $\Delta = \Delta(L)$ be the commuting graph of L. Our notation for graphs and the commuting graph are as introduced in Section 1 (see 1.3), in particular, for $a \in \Delta$, $\Delta^i(a) = \{x \in \Delta : d(a, x) = i\}$ (d is the distance function) and $\Delta(a) = \Delta^1(a)$.

The purpose of this section is to prove that Δ is balanced (Theorem 9.14), we do this by showing that, in the notation of 9.2 (below), there exists $a \in \Delta$ such that $\Xi(a) \neq \emptyset$. Then, by definition, for each $b \in \Xi(a)$, $B_\Delta(a, b)$ and $B_\Delta(b, a)$, so Δ is balanced.

Notation. We denote $SL_n^\epsilon(q) = SL_n(q), SU_n(q)$, according to whether $\epsilon = 1, -1$. Similarly for GL_n^ϵ and PSL_n^ϵ.

In what follows we take $\epsilon = 1$, unless $4 \mid q - 1$, in which case we take $\epsilon = -1$. Of course $4 \nmid q - \epsilon$.

9.1. (1) *L contains a subgroup $K \cong PSL_8^\epsilon(q)$.*

(2) *K contains a subgroup $H \cong GL_7^\epsilon(q)/\mathbb{Z}_{(2,q-\epsilon)}$, which contains a cyclic maximal torus of order $(q^7-\epsilon)/(2,q-\epsilon)$.*
(3) *$Z(H) \cong \mathbb{Z}_{(q-\epsilon)/(2,q-\epsilon)}$, a group of odd order.*
(4) *Let $1 \neq a \in Z(H)$. Then $C_L(a) = H$.*

Proof. View $L = (\bar{L}_\sigma)'$, where $\bar{L}$ is an adjoint group of type E_7 and σ is a Frobenius morphism. Then L has index δ in $\bar{L}_\sigma$. There is a σ-invariant maximal rank subgroup $A_7 < \bar{L}$ with center of order δ. Then $N_{E_7}(A_7) = A_7.2$, the extra involution being the long word in a suitable Weyl group and inducing a graph automorphism on A_7. It follows from [**1**, I, 2.8], that there are two classes of σ-invariant conjugates of A_7. For elements in one class σ induces a field morphism and on the other a graph-field morphism. Let $\overline{E}$ be an element of one of these classes, determined by ϵ. Then $\overline{E}_\sigma < \bar{L}_\sigma$.

Let $\hat{E} = SL_8$, the simply connected group of type A_7. There is a surjective homomorphism $\theta : \hat{E} \to \overline{E}$, with kernel of order 4 or 1, according to whether q is odd or even. Moreover, there is a Frobenius morphism of $\hat{E}$, which we also call σ, which commutes with θ.

Now $\hat{K} = (\hat{E})_\sigma = SL_8^\epsilon(q)$ and this group contains $\hat{H} \cong GL_7^\epsilon(q)$, which arises by taking fixed points of a σ-invariant subgroup of $\hat{E}$ of type A_6T_1.

Set $K = \theta(\hat{K})$, so that $K \cong SL_8^\epsilon(q)/\mathbb{Z}_{(4,q-\epsilon)}$. Our choice of ϵ forces $K \cong PSL_8^\epsilon(q)$ giving (1).

Let $\overline{D} = \theta(A_6T_1) < \overline{E}$. Then $\overline{D}_\sigma$ and $(A_6T_1)_\sigma$ have the same order (see the proof of (2.12) in [**15**]), so $\overline{D}_\sigma \geq \theta(GL_7^\epsilon(q))$ as a subgroup of index $(4, q-\epsilon)$. Also $\overline{D}_\sigma$ covers $\bar{L}_\sigma/L$.

Our choice of ϵ implies that $GL_7^\epsilon(q) = J \times S$, where $J = O^{2'}(GL_7^\epsilon(q))$ and $S \cong \mathbb{Z}_{(2,q-\epsilon)}$. Then θ restricts to an isomorphism on J and setting $H = \theta(J)$ we obtain (2). We note that H has index $(2, q-\epsilon)$ in $\overline{D}_\sigma$, and if the index is 2, then there is an involution in $\overline{D}_\sigma$ which is in $\bar{L}_\sigma - L$ ((2.12) in [**15**]). Also H contains a cyclic maximal torus of order $(q^7-\epsilon)/(2,q-\epsilon)$. Thus (2) holds. (3) follows from (2) and our choice of ϵ.

Fix $1 \neq a \in Z(H)$. Then $C_{\overline{L}}(a)^0 \geq \overline{D}$, a maximal rank group of type A_6T_1. If the containment is strict, then $C_{\overline{L}}(a)^0$ would have to be a semisimple group of rank 7. But a consideration of root systems shows that the only such subgroups of E_7 containing A_6 are of type A_7 and such a group has centralizer of order at most 2. Thus equality holds and taking fixed points we have $C_{\overline{L}_\sigma}(a) = \overline{D}_\sigma$. Intersecting with L yields (4).

9.2. Notation and definitions.

(1) $\mathcal{T}$ denotes the set of maximal tori in L of order $(q^7-\epsilon)/(2,q-\epsilon)$ as in 9.1.2. Of course $\mathcal{T}$ is a conjugacy class of tori in L.
(2) Given $T \in \mathcal{T}$, we denote by $R_T \leq T$, the unique subtorus of order $(q-\epsilon)/(2,q-\epsilon)$. We let $\Lambda_T = T - R_T$. We set $\Lambda = \cup_{T\in\mathcal{T}}\Lambda_T$ and we let $\lambda = |\Lambda|$.

(3) Given $T \in \mathcal{T}$, we let $H_T = C_L(R_T)$.
Let $a \in \Lambda$.
(4) We let $\Theta(a) = \Delta^{\leq 3}(a)$. We denote $\theta = |\Theta(a)|$. We'll see in 9.3 below that θ is independent of a.
(5) We let $\Gamma(a) = \{b \in \Lambda : d(a, ab) > 3 < d(a, a^{-1}b)\}$.
(6) We denote $\Gamma^*(a) = \{b \in \Lambda : a \in \Gamma(b)\}$.
(7) We denote $\Xi(a) = \Gamma(a) \cap \Gamma^*(a) \cap \Lambda^{>3}(a)$.

9.3. *Let $a \in \Lambda$. Then:*
(1) *There exists a unique $T \in \mathcal{T}$ such that $a \in T$. Further, $C_L(a) = T$.*
Let $T \in \mathcal{T}$ be the unique torus containing $\{a\}$. Then:
(2) $\Delta(a) = T - \{1, a\}$.
(3) $\Delta^2(a) = H_T - T$.
(4) $|\Delta^k(a)| = |\Delta^k(b)|$, *for all $b \in \Lambda$ and all k.*

Proof. Let $a \in \Lambda$. To show (1), suppose first that the order of a, $|a|$ is not a power of 7. We claim that a satisfies the hypotheses for s in 8.1. Recall that if $S_\sigma \leq \tilde{L}$ is a maximal torus, then $|S_\sigma| = g(q)$, where $g(x)$ is a polynomial of degree 7, a product of cyclotomic polynomials, hence the hypotheses of 8.1 follow from 8.6.5 if $\epsilon = -1$ and from 8.6.6, if $\epsilon = 1$. So suppose $|a|$ is a power of 7. Let $T \in \mathcal{T}$ such that $a \in T$. Since T is cyclic, $1 \neq a^k \in R_T$, for some $k \geq 2$. Then $C_L(a) \leq C_L(a^k) = C_L(R_T)$, by 9.1.4. Hence, (1) follows from inspecting $C_H(a)$, where $H = H_T$. This shows (1). Now, (2) is immediate from (1), and (3) is immediate from (2) and 9.1.4. Also (3) says that $\Delta^2(x) = \Delta^2(y)$, for $x, y \in \Lambda_T$, so since $\mathcal{T}$ is a conjugacy class of subgroups, (4) follows.

9.4. *Let $a \in \Lambda$ and set $\Theta = \Theta(a)$. Then:*
(1) $\Gamma(a) = \Lambda - ((a^{-1}(a\Lambda \cap \Theta)) \cup (a(a^{-1}\Lambda \cap \Theta)))$.
(2) $|\Gamma(a)| \geq \lambda - 2\theta$.

Proof. Note that $\{b \in \Lambda : d(a, ab) \leq 3\} = a^{-1}(a\Lambda \cap \Theta(a))$ and $\{b \in \Lambda : d(a, a^{-1}b) \leq 3\} = a(a^{-1}\Lambda \cap \Theta(a))$. Hence (1) holds. (2) is immediate from (1).

9.5. *There exists $a \in \Lambda$ such that $|\Gamma^*(a)| \geq \lambda - 2\theta$.*

Proof. Let $M = \mathrm{Max}_{b\in\Lambda}|\Gamma^*(b)|$. Count the number of pairs $X = \{(a, b) : a, b \in \Lambda \text{ and } b \in \Gamma(a)\}$. Using 9.4, we have $\lambda(\lambda - 2\theta) \leq \sum_{a\in\Lambda} |\Gamma(a)| = |X| = \sum_{b\in\Lambda} |\Gamma^*(b)| \leq \lambda M$. Thus $M \geq (\lambda - 2\theta)$ as asserted.

9.6. Notation. From now on we fix $a \in \Lambda$ such that $|\Gamma^*(a)| \geq \lambda - 2\theta$, and we set $\Theta = \Theta(a)$, $\Xi = \Xi(a)$ and $\xi = |\Xi|$. Let T denote the unique member of $\mathcal{T}$ containing $\{a\}$ and set $H = H_T$.

9.7. (1) $|\Gamma(a) \cap \Gamma^*(a)| \geq \lambda - 4\theta$.
(2) $\xi \geq \lambda - 5\theta$.

Proof. $|\Gamma(a)\cap\Gamma^*(a)| \geq |\Gamma(a)|-|\Lambda-\Gamma^*(a)| \geq (\lambda-2\theta)-(\lambda-(\lambda-2\theta)) = \lambda-4\theta$. The proof of (2) is similar.

The remainder of this section is devoted to showing that $\Xi \neq \emptyset$, or that $\xi > 0$. It will be done by producing an upper bound to θ. To estimate sizes of subgroups we'll use the following lemma.

9.8. *Let* $2 \leq a_1 < a_2 < \ldots < a_k$ *be integers and let* $\epsilon_1, \epsilon_2, \ldots, \epsilon_k \in \{1, -1\}$. *Then*

$$\frac{1}{2} \leq \frac{(q^{a_1}+\epsilon_1)(q^{a_2}+\epsilon_2)\cdots(q^{a_k}+\epsilon_k)}{q^{a_1+a_2+\cdots+a_k}} \leq 2.$$

Proof. This is taken from [**18**, p. 2100]. We include the proof in [**18**]. For $i \geq 2$, we have

$$1-\frac{1}{2^i} \geq \frac{\frac{1}{2}+\frac{1}{2^i}}{\frac{1}{2}+\frac{1}{2^{i-1}}}, \quad 1+\frac{1}{2^i} \leq \frac{1-\frac{1}{2^i}}{1-\frac{1}{2^{i-1}}}.$$

Therefore the fraction

$$\frac{(q^{a_1}+\epsilon_1)(q^{a_2}+\epsilon_2)\cdots(q^{a_k}+\epsilon_k)}{q^{a_1+a_2+\cdots+a_k}}$$

is at least

$$\prod_{i=1}^{k}\left(1-\frac{1}{q^{a_i}}\right) \geq \prod_{i=2}^{K}\left(1-\frac{1}{q^{i}}\right) \geq \prod_{i=2}^{K}\left(1-\frac{1}{2^{i}}\right)$$
$$\geq \prod_{i=2}^{K}\frac{\frac{1}{2}+\frac{1}{2^i}}{\frac{1}{2}+\frac{1}{2^{i-1}}} = \frac{1}{2}+\frac{1}{2^K} > \frac{1}{2},$$

(where $K = a_k$) and at most

$$\prod_{i=1}^{k}\left(1+\frac{1}{q^{a_i}}\right) \leq \prod_{i=2}^{K}\left(1+\frac{1}{q^{i}}\right) \leq \prod_{i=2}^{K}\left(1+\frac{1}{2^{i}}\right)$$
$$\leq \prod_{i=2}^{K}\frac{1-\frac{1}{2^i}}{1-\frac{1}{2^{i-1}}} = 2-\frac{1}{2^{K-1}} < 2.$$

9.9. (1) $|H| \leq 3q^{49} \leq q^{51}$.
(2) $|L| \geq \frac{1}{2\delta}q^{133}$.
(3) $\lambda \geq \frac{1}{14\delta}q^{133}$.

Proof. By 9.1.2, $|H| = \frac{1}{(2,q-\epsilon)}\,|GL_7^\epsilon(q)|$. By 9.8, $|SL_7^\epsilon(q)| \leq 2q^{48}$. Hence, $\frac{1}{(2,q-\epsilon)}|GL_7^\epsilon(q)| = \frac{1}{(2,q-\epsilon)}(q-\epsilon)|SL_7^\epsilon(q)| \leq \frac{2}{(2,q-\epsilon)}(q-\epsilon)q^{48} \leq 3q^{49}$. (2) follows immediately from 9.8. Now $|\Lambda_T| = |T-R_T| = \frac{1}{(2,q-\epsilon)}\{q^7-\epsilon-(q-\epsilon)\} =$

$\frac{1}{(2,q-\epsilon)}(q^7-q)$. Since every element of Λ lies in a unique member of $\mathcal{T}$, we get that

$$|\Lambda| = |\Lambda_T|\,|\mathcal{T}| \geq \frac{1}{(2,q-\epsilon)}(q^7-q)\cdot\frac{|L|}{7|T|} = \frac{1}{7\delta}|L|\,\frac{q^7-q}{q^7-\epsilon}$$
$$= \frac{1}{7\delta}q^{63}(q^7-q)(q^7+\epsilon)(q^{18}-1)(q^{12}-1)(q^{10}-1)(q^8-1)(q^6-1)(q^2-1)$$
$$\geq \frac{1}{14\delta}q^{133}$$

by 9.8; notice that the argument in the proof of 9.8 applies even though we have q^6-1 appearing twice in the last product.

Corollary 9.10. (1) *Suppose* $\theta < \frac{1}{70\delta}q^{133}$. *Then* $\xi > 0$.
(2) *Suppose* $\theta < q^{126}$. *Then* $\xi > 0$.

Proof. By 9.7.2, $\xi \geq \lambda - 5\theta$. Now $\lambda - 5\theta > 0$, iff $\lambda > 5\theta$ iff $\theta < \frac{1}{5}\lambda$. By 9.9.3, $\lambda \geq \frac{1}{14\delta}q^{133}$, so $\frac{1}{5}\lambda \geq \frac{1}{70\delta}q^{133}$. (2) follows immediately from (1).

9.11. *Let* $\mathfrak{M} = \{h \in H - \{1\} : |C_L(h)| \geq q^{74}\}$. *Set* $\mathbb{M} = \cup_{h\in\mathfrak{M}} C_L(h)$ *and* $\mu = |\mathbb{M}|$. *If* $\mu \leq q^{125}$, *then* $\xi > 0$.

Proof. By 9.10.2, it suffices to show that $\theta \leq q^{126}$. Of course, by 9.3.3, any element in Θ centralizes a nontrivial element of H. Hence

$$\text{(i)} \qquad \theta \leq \left|\bigcup_{h\in H-\{1\}} C_L(h)\right|.$$

Let $\mathbb{M}_1 = \bigcup\{C_L(h) : 1 \neq h \in H - \mathfrak{M}\}$. Of course, $|\mathbb{M}_1| \leq \sum_{1\neq h\in H-\mathfrak{M}} |C_L(h)| < |H|q^{74} \leq q^{125}$. Also, $\bigcup\{C_L(h) : 1 \neq h \in H\} = \mathbb{M}_1 \cup \mathbb{M}$, so by (i), $\theta \leq |\mathbb{M}_1| + |\mathbb{M}| \leq q^{125} + q^{125} \leq q^{126}$.

Hence, it remains to show that $\mu \leq q^{125}$.

9.12. *Let* $x \in H$ *satisfy* $|C_L(x)| \geq q^{74}$. *Then one of the following holds:*

(1) *x is unipotent of class A_1, $|C_L(x)| \leq 2q^{99}$, $x^L \cap H$ is a conjugacy class of H and $|H : C_H(x)| \leq 4q^{12}$.*
(2) *x is unipotent of class $2A_1$, $|C_L(x)| \leq 2q^{81}$, $x^L \cap H$ is a conjugacy class of H, and $|H : C_H(x)| \leq 4q^{20}$.*
(3) *x is semisimple, $C_L(x)' \cong E_6(q)$ or ${}^2E_6(q)$ according to whether $\epsilon = 1$ or -1. $C_L(x) = C_L(x)'S$, where S is cyclic of order $(q-\epsilon)/(2, q-\epsilon)$. Hence $|C_L(x)| \leq 3q^{79}$. Either $|H : C_H(x)| = |GL_7^\epsilon(q) : GL_5^\epsilon(q)GL_2^\epsilon(q)| \leq 4q^{20}$ or $|H : C_H(x)| = |GL_7^\epsilon(q) : GL_6^\epsilon(q)GL_1^\epsilon(q)| \leq 2q^{12}$.*

Proof. Write $x = su$ as a commuting product of a semisimple and a unipotent element. Then $C_L(x) \leq C_L(s)$. The latter group is obtained by taking the set of fixed points under σ from the centralizer in the algebraic group, then intersecting with L. In the algebraic group the centralizer is a reductive subgroup of maximal rank and a trivial check of subsystems shows that the only subsystems giving a large enough centralizer are of type E_7 or E_6T_1. In the first case, $s = 1$ and in the latter case $u = 1$ in order to have large enough centralizer (see [**7**]).

Suppose $s = 1$, so that x is unipotent. Then a check of [**8**] shows that x has types A_1, $(2A_1)$, or $(3A_1)''$. Now x is contained in a subsystem subgroup of $\tilde{L}$ of type A_6. The Jordan form of a unipotent element of A_6 determines a subsystem group containing the unipotent element as a regular element. Each of the relevant subsystems is a Levi factor, so by the classification of unipotent elements, x must also be of type $A_1, 2A_1$, or $3A_1$ within A_6.

Now E_7 has just one class of subsystem groups of type A_1 and $2A_1$, but it has two classes of subsystem groups of type $3A_1$ and we claim that the class $(3A_1)''$ is not represented in A_6. To see this start from a subsystem group of type A_1, with centralizer D_6. Working in A_1D_6 we see that there are two classes of groups of type $3A_1$, with centralizers $D_4, 4A_1$, respectively. Only unipotent elements of type $(3A_1)''$ have centralizer involving D_4, so the former class is of type $(3A_1)''$. On the other hand, the group $3A_1$ in A_6 is contained in A_1A_4, so from the centralizer of the first factor we get $A_4 < D_6$ and from here we see that the full centralizer of $3A_1$ cannot contain D_4, so this must be the class $(3A_1)'$, establishing the claim.

One checks that the centralizers of unipotent elements of type A_1 and $2A_1$ in $A_6T_1 \cong GL_7$ are connected, so each type is represented by a single class in $GL_7^{\epsilon}(q)$ ([**1**, I, 2.8]) and hence in H. Centralizers are given in [**8**], so the numerical information in (1) and (2) follows by taking fixed points and using 9.8.

Now suppose $s \neq 1$. We again consider the group $A_7 = \overline{E} < \bar{L}$. It is shown in (2.3) of [**6**] the the 56-dimensional restricted module for a simple connected group of type E_7 restricts to a subgroup of type A_7 as the wedge square of the natural module and its dual. In each of these three irreducible modules the Weyl group of E_7 or A_7 with respect to a maximal torus is transitive on weight spaces within the module. The stabilizer in $W(E_7)$ of a weight space is $W(E_6)$ and this is also the centralizer in $W(E_7)$ of the central torus in $C_L(s)$.

Choose a σ-invariant maximal torus $R < \overline{E}$. Taking Weyl groups with respect to R, it follows from the above paragraph that $W(A_7)$ has two orbits on 1-dimensional tori in R, with centralizer of type $W(E_6)$. Each has stabilizer in $W(A_7)$ of type $W(A_5)W(A_1)$. So for such a 1-dimensional torus, the centralizer in A_7 is a reductive group with Weyl group of type $W(A_5)W(A_1)$. The only possibility is that the centralizer has the form $A_5A_1T_1$.

Elements of the above 1-dimensional torus are represented in $\overline{E}$ as images of elements of SL_8 having one eigenvalue of multiplicity 6 and another of multiplicity 2. Taking fixed points and working in $GL_7^{\epsilon}(q)$ we see that there are two types of semisimple elements in H of the correct type. In the action on the natural 7-dimensional module one type has one eigenvalue of multiplicity 6 and one eigenvalue of multiplicity 1, while for the other class there is one eigenvalue of multiplicity 5 and another of multiplicity 2. The conclusion follows.

Corollary 9.13. $\mu \leq q^{125}$.

Proof. For $i = 1, 2$ let u_i denote a unipotent element as in 9.12.1, and set $M_i = u_i^L \cap H$. Let S_1, S_2 be subgroups of order $(q-\epsilon)/(2, q-\epsilon)$ in H corresponding to subgroups of $GL_7^\epsilon(q)$ with centralizer $GL_5^\epsilon(q)GL_2^\epsilon(q)$ or $GL_6^\epsilon(q)GL_1^\epsilon(q)$, respectively. We claim that $C_L(S) = C_L(y)$, for $S \in \{S_1, S_2\}$ and $1 \neq y \in S$. This follows from the fact that the preimage of S in $\tilde{L}$ has centralizer of type E_6T_1, which is maximal among reductive subgroups of E_7. Recall that we defined $\mathfrak{M} = \{h \in H - \{1\} : |C_L(h)| \geq q^{74}\}$ and $\mathbb{M} = \bigcup_{h \in \mathfrak{M}} C_L(h)$. By 9.12 we have

$$\mu = |\mathbb{M}| \leq \sum_{x \in M_1} |C_L(x)| + \sum_{x \in M_2} |C_L(x)| + \sum_{x \in S_1^H} |C_L(x)| + \sum_{x \in S_2^H} |C_L(x)|.$$

Hence $\mu \leq (2q^{99})(4q^{12}) + (2q^{81})(4q^{20}) + (3q^{79})(4q^{20}) + (3q^{79})(4q^{12}) \leq q^{125}$.

Theorem 9.14. *Δ is balanced.*

Proof. By 9.13, $\mu \leq q^{125}$, so by 9.11 $\xi > 0$. Hence $\Xi(a) \neq \emptyset$ and as we remarked at the beginning of Section 9, this shows (by definition) that Δ is balanced.

10. The Alternating Groups.

In this section A_m denote the Alternating Group on $\{1, 2, \dots, m\}$. The purpose of this section is to prove the following theorem:

Theorem 10.1. *Let $m > 3$ and let $L \cong A_m$. Then* $\operatorname{diam}(\Delta(L)) > 4$.

Throughout this section $n > 2$ is a fixed even integer, such that $n-1$ is not a prime. We let G be the Symmetric Group on $\{1, 2, \dots, n\}$. We use cyclic notation for permutations in G. We apply permutations on the right, so for $\sigma \in G$, and $i \in \{1, 2, \dots, n\}$, $i\sigma$ is the image of i under σ. In addition, when we write a permutation as a product of cycles, the even numbers that occur are bolded and enlarged. For example, if $1 \leq k \leq n$ is an odd number congruent to 1 (mod 4), then

$$\rho = (1, 5, 9, \cdots, k)(\mathbf{k+1}, \mathbf{k+3}, \cdots, \mathbf{2k})$$

is the permutation with $i\rho = i+4$, if $1 \leq i \leq k-4$ is congruent to 1 (mod 4), $k\rho = 1$, $i\rho = i+2$, if $k+1 \leq i \leq 2k-2$ is even, and $(2k)\rho = k+1$.

Another convention that we'll use is that $\cdots$ means continue with the same pattern. Thus for example, in ρ, the $\cdots$ after 9 means that $9\rho = 13$, $13\rho = 17$, and so on until we get to $k-4$. Another example is

$$\eta = (1, \mathbf{2}, 3, \cdots, \mathbf{k-1}, \mathbf{k+3}, \cdots, \mathbf{4k})$$

is a cycle such that $i\eta = i+1$, $1 \le i \le k-2$, $i\eta = i+4$, if $k-1 \le i \le 4k-4$, is congruent to 0 (mod 4) and $(4k)\eta = 1$.

Notation. (1) For a permutation $\sigma \in G$, we denote by $\operatorname{supp}(\sigma)$ the set of elements moved by σ.

(2) We fix once and for all the letter g to denote the permutation

$$g = g_n = (1, \mathbf{2}, 3, \cdots, \mathbf{n-2}, n-1).$$

(3) We fix once and for all the letter s to denote the permutation

$$s = s_n = (3, \mathbf{4})(5, \mathbf{6}) \cdots (n-1, \mathbf{n}).$$

(4) Let p be a prime divisor of $n-1$. We write $n_p = \frac{n-1}{p}$. Thus $n-1 = pn_p$.

(5) Let p be a prime divisor of $n-1$. We denote

$$\theta_p = g^{n_p}.$$

The main result of this section, from which Theorem 10.1 follows, is the following theorem.

Theorem 10.2. *Let $n > 2$ be an even number. Suppose $n-1$ is not a prime and let p, q be prime divisors of $n-1$, with $p \le q$. Let $\Gamma = \langle \theta_p, s\theta_q^{-1}s \rangle$. Then:*

(1) *Γ is a transitive subgroup of G.*

(2) $C_G(\Gamma) = \{1\}$.

We'll now prove Theorem 10.1, under the assumption that Theorem 10.2 holds.

Proof of Theorem 10.1. Let $L = A_m$. We assume that Theorem 10.2 holds and we prove Theorem 10.1. Let d be the distance function on $\Delta(L)$. Suppose first that m is even. If $m-1$ is a prime, then it is easy to check that $\langle g_m \rangle - \{1\}$ is a connected component of $\Delta(L)$. So assume $m-1$ is a composite odd number. Let $g = g_m$ and $s = s_m$. We'll show that $d(g, sg^{-1}s) > 4$. So suppose $d(g, sg^{-1}s) \le 4$. Since $C_L(g) = \langle g \rangle$, and $C_L(sg^{-1}s) = \langle sg^{-1}s \rangle$, there are prime divisors p, q of $m-1$ such that $\pi := g, g^{\frac{(m-1)}{p}}, x, sg^{\frac{(1-m)}{q}}s, sg^{-1}s$ is a path in $\Delta(L)$. But then $x \in C_L\left(\left\langle g^{\frac{(m-1)}{p}}, sg^{\frac{(1-m)}{q}}s \right\rangle\right)$, so if $p \le q$, this contradicts Theorem 10.2, while if $p > q$, then inverting the path π and conjugating by s, we get that $g, g^{\frac{(m-1)}{q}}, sx^{-1}s, sg^{\frac{(1-m)}{p}}s, sg^{-1}s$ is also a path in $\Delta(L)$, and this contradicts Theorem 10.2.

Suppose next that m is odd. If $m-2$ is a prime, then $\langle g_{m-1} \rangle - \{1\}$ is a connected component of $\Delta(L)$. So assume $m-2$ is a composite odd number. Let $g = g_{m-1}$ and $s = s_{m-1}$. Let p, q be prime divisors of $m-2$. Let $\Gamma = \left\langle g^{\frac{(m-2)}{p}}, sg^{\frac{(2-m)}{q}}s \right\rangle$. By Theorem 10.2, $\{1, 2, \cdots, m-1\}$ is an orbit of Γ, so the centralizer of Γ in L fixes m, and hence by Theorem 10.2, it is trivial. Then, the same proof as in the case when m is even shows that $d(g, sg^{-1}s) > 4$.

10.3. *Let p be a prime divisor of $n-1$. Then:*

(1) $\theta_p = (1, \mathbf{n}_p + \mathbf{1}, \cdots, (p-1)n_p + 1)(\mathbf{2}, n_p + 2, \cdots, (\mathbf{p}-\mathbf{1})\mathbf{n}_p + \mathbf{2}) \cdots (n_p, \mathbf{2n}_p, \cdots, n-1)$ *and* θ_p *fixes* n.

(2) *Two indices $i, j \in \{1, 2, \cdots, n-1\}$ are in the same orbit of θ_p, iff they are congruent modulo n_p.*

(3) *For all $1 \leq i \leq n-1$, and all integers $k, ig^k = k + i$, in particular, $i\theta_p = n_p + i$, and $i\theta_p^{-1} = i - n_p$, where indices are taken modulo $(n-1)$.*

(4) *For $\sigma \in G$, and $i, k \in \{1, \ldots, n-1\}$, if $i\sigma = j \neq n$, then $(k+i)g^{-k}\sigma g^k = k+j$ and $(i-k)g^k\sigma g^{-k} = j-k$, in particular, $(n_p+i)\theta_p^{-1}\sigma\theta_p = n_p+j$, $(i-n_p)\theta_p\sigma\theta_p^{-1} = j-n_p$ and $(n_p-n_q+i)g^{(n_q-n_p)}\sigma g^{(n_p-n_q)} = n_p-n_q+j$, where indices are taken modulo $(n-1)$.*

Proof. The proof is straightforward.

Important Remark. In order to verify the calculations in this section, we emphasize that n_p denotes $\mathbf{\frac{n-1}{p}}$ and not $\frac{n}{p}$. In addition $ig^k = i + k$, **modulo (n − 1)** and not modulo n.

Notation. From now on we fix two primes p and q dividing $n-1$, such that $p \leq q$.

10.4.

(1) $\theta_q^{-1}s\theta_q = (\mathbf{n}_q + \mathbf{3}, n_q + 4)(\mathbf{n}_q + \mathbf{5}, n_q + 6) \cdots (\mathbf{n} - \mathbf{2}, n-1)(1, \mathbf{2})(3, \mathbf{4}) \cdots (n_q - 2, \mathbf{n}_q - \mathbf{1})(n_q, \mathbf{n})$.

(2) $\theta_q s\theta_q^{-1} = (n - n_q + 2, \mathbf{n} - \mathbf{n}_q + \mathbf{3}) \cdots (n-3, \mathbf{n} - \mathbf{2})(n-1, 1)(\mathbf{2}, 3)(\mathbf{4}, 5) \cdots (\mathbf{n} - \mathbf{n}_q - \mathbf{3}, n - n_q - 2)(\mathbf{n} - \mathbf{n}_q - \mathbf{1}, \mathbf{n})$.

(3) $\theta_q^{-1}s\theta_q s = (n_q, n-1, n-3, \cdots, n_q+4, n_q+2, \mathbf{n}_q + \mathbf{3}, \mathbf{n}_q + \mathbf{5}, \cdots, \mathbf{n} - \mathbf{2}, \mathbf{n}, \mathbf{n}_q + \mathbf{1})(1, \mathbf{2})$.

(4) $\theta_q s\theta_q^{-1}s = (\mathbf{2}, \mathbf{4}, \cdots, \mathbf{n} - \mathbf{n}_q - \mathbf{1}, n-1, 1, \mathbf{n}, n - n_q - 2, n - n_q - 4, \cdots, 3)(n - n_q, \mathbf{n} - \mathbf{n}_q + \mathbf{1})$.

(5) $[\theta_p, s\theta_q^{-1}s] = g^{(n_q-n_p)}\theta_q^{-1}s\theta_q sg^{(n_p-n_q)}\theta_q s\theta_q^{-1}s$.

(6) *If $p \neq q$, then*

$$g^{(n_q-n_p)}\theta_q^{-1}s\theta_q sg^{(n_p-n_q)} =$$

$$(n_p, \mathbf{n}_p - \mathbf{n}_q, \mathbf{n}_p - \mathbf{n}_q - \mathbf{2}, \cdots, \mathbf{2}, n-1, n-3, \cdots, n_p + 2, \mathbf{n}_p + \mathbf{3}, \mathbf{n}_p + \mathbf{5}, \cdots, \mathbf{n} - \mathbf{2}, 1, 3, \cdots, n_p - n_q - 1, \mathbf{n}, \mathbf{n}_p + \mathbf{1})(n_p - n_q + 1, \mathbf{n}_p - \mathbf{n}_q + \mathbf{2}).$$

Proof. For (1), we have,
$\theta_q^{-1}s\theta_q = (3\theta_q, 4\theta_q)(5\theta_q, 6\theta_q)\cdots((n-1)\theta_q, n\theta_q) =$

$$(\mathbf{n}_q + \mathbf{3}, n_q + 4)(\mathbf{n}_q + \mathbf{5}, n_q + 6)\cdots(\mathbf{n} - \mathbf{2}, n-1)(1, \mathbf{2})(3, \mathbf{4})$$
$$\cdots(n_q - 2, \mathbf{n}_q - 1)(n_q, \mathbf{n})$$

where we use 10.3 to verify this equality, noting that θ_q fixes n. (2) is proved similarly.

We now prove (3). We first write $\theta_q^{-1}s\theta_q$ and s one below the other.

$$(\mathbf{n}_q + \mathbf{3}, n_q + 4)(\mathbf{n}_q + \mathbf{5}, n_q + 6)\cdots(\mathbf{n} - \mathbf{2}, n-1)(1, \mathbf{2})(3, \mathbf{4})$$
$$\cdots(n_q - 2, \mathbf{n}_q - \mathbf{1})(n_q, \mathbf{n})\cdot$$
$$(3, \mathbf{4})(5, \mathbf{6})\cdots(n-3, \mathbf{n} - \mathbf{2})(n-1, \mathbf{n}) = .$$

Note that $(3, \mathbf{4})(5, \mathbf{6})\cdots(n_q - 2, \mathbf{n}_q - \mathbf{1})$ is canceled. Hence

$$= (\mathbf{n}_q + \mathbf{3}, n_q + 4)(\mathbf{n}_q + \mathbf{5}, n_q + 6)\cdots(\mathbf{n} - \mathbf{2}, n-1)(1, \mathbf{2})(n_q, \mathbf{n})\cdot$$
$$(n_q, \mathbf{n}_q + \mathbf{1})(n_q + 2, \mathbf{n}_q + \mathbf{3})\cdots(n-3)(\mathbf{n} - \mathbf{2})(n-1, \mathbf{n}) = .$$

Now start with n_q and carefully work though the product.

$$= (n_q, n-1, n-3, \cdots, n_q + 4, n_q + 2, \mathbf{n}_q + \mathbf{3}, \mathbf{n}_q + \mathbf{5}, \cdots,$$
$$\mathbf{n} - \mathbf{2}, \mathbf{n}, \mathbf{n}_q + 1)(1, \mathbf{2}).$$

Next we prove (4). We first write $\theta_q s\theta_q^{-1}$ and s one below the other.

$$(n - n_q + 2, \mathbf{n} - \mathbf{n}_q + \mathbf{3})\cdots(n-3, \mathbf{n} - \mathbf{2})(n-1, 1)(\mathbf{2}, 3)(\mathbf{4}, 5)\cdots$$
$$(\mathbf{n} - \mathbf{n}_q - \mathbf{3}, n - n_q - 2)(\mathbf{n} - \mathbf{n}_q - \mathbf{1}, \mathbf{n})\cdot$$
$$(3, \mathbf{4})(5, \mathbf{6})\cdots(n-3, \mathbf{n} - \mathbf{2})(n-1, \mathbf{n}) = .$$

Note that $(n - n_q + 2, \mathbf{n} - \mathbf{n}_q + \mathbf{3})\cdots(n-3, \mathbf{n} - \mathbf{2})$ is canceled. Hence

$$= (n-1, 1)(\mathbf{2}, 3)(\mathbf{4}, 5)\cdots(\mathbf{n} - \mathbf{n}_q - \mathbf{3}, n - n_q - 2)(\mathbf{n} - \mathbf{n}_q - \mathbf{1}, \mathbf{n})$$
$$(3, \mathbf{4})(5, \mathbf{6})\cdots(n - n_q - 2, \mathbf{n} - \mathbf{n}_q - \mathbf{1})(n - n_q, \mathbf{n} - \mathbf{n}_q + \mathbf{1})(n-1, \mathbf{n})$$
$$= (\mathbf{2}, \mathbf{4}, \cdots, \mathbf{n} - \mathbf{n}_q - \mathbf{1}, n-1, 1, \mathbf{n}, n - n_q - 2, n - n_q - 4,$$
$$\cdots, 3)(n - n_q, \mathbf{n} - \mathbf{n}_q + \mathbf{1}).$$

We now compute $[\theta_p, s\theta_q^{-1}s] = \theta_p^{-1}s\theta_q s\theta_p s\theta_q^{-1}s$. Recall that by definition, $\theta_p = g^{n_p}$ and $\theta_q = g^{n_q}$. Hence $[\theta_p, s\theta_q^{-1}s] = g^{(n_q - n_p)}\theta_q^{-1}s\theta_q s g^{(n_p - n_q)}\theta_q s\theta_q^{-1}s$.

Finally,

$$\begin{gathered} g^{(n_q-n_p)}\theta_q^{-1}s\theta_q s g^{(n_p-n_q)} \\ = g^{(n_q-n_p)}(n_q, n-1, n-3, \cdots, n_q+4, n_q+2, \mathbf{n}_q+\mathbf{3}, \mathbf{n}_q+\mathbf{5}, \\ \cdots, \mathbf{n}-\mathbf{2}, \mathbf{n}, \mathbf{n}_q+\mathbf{1})(1,\mathbf{2})g^{(n_p-n_q)}. \end{gathered}$$

Now using 10.3.4 we get

$$\begin{gathered} = (n_p, \mathbf{n}_p-\mathbf{n}_q, \mathbf{n}_p-\mathbf{n}_q-\mathbf{2}, \cdots \mathbf{2}, n-1, n-3, \\ \cdots, n_p+2, \mathbf{n}_p+\mathbf{3}, \mathbf{n}_p+\mathbf{5}, \cdots, \mathbf{n}-\mathbf{2}, \\ 1, 3, \cdots, n_p-n_q-1, \mathbf{n}, \mathbf{n}_p+\mathbf{1})(n_p-n_q+1, \mathbf{n}_p-\mathbf{n}_q+\mathbf{2}). \end{gathered}$$

10.5. *Suppose $n_p - n_q > 2$, then:*

(1) *The fixed points of $[\theta_p, s\theta_q^{-1}s]$ are*

$$\{3, \mathbf{4}, \ldots, n_p-n_q-3, \mathbf{n}_p-\mathbf{n}_q-\mathbf{2}, \mathbf{n}_p-\mathbf{n}_q\}$$

where if $n_p - n_q = 4$, then $\{4\}$ is the unique fixed point.

(2) *If $n - n_p - n_q \equiv 2 \pmod 4$, then $[\theta_p, s\theta_q^{-1}s] =$*

$$\begin{gathered} (1, \mathbf{2}) \cdot \\ (n_p-n_q-1, n-n_q-2, n-n_q-6, \cdots, n_p, \mathbf{n}_p-\mathbf{n}_q+\mathbf{2}) \cdot \\ (n_p-2, n_p-4, \cdots, n_p-n_q+1, \mathbf{n}_p-\mathbf{n}_q+\mathbf{4}, \\ \mathbf{n}_p-\mathbf{n}_q+\mathbf{6}, \cdots, \mathbf{n}_p-\mathbf{1}, \mathbf{n}_p+\mathbf{1}) \cdot \\ (n-n_q, n-n_q-4, \cdots, n_p+4, n_p+2, \mathbf{n}_p+\mathbf{5}, \mathbf{n}_p+\mathbf{9}, \cdots, \mathbf{n}-\mathbf{n}_q-\mathbf{1}) \cdot \\ (n-1, n-3, \cdots, n-n_q+2, \mathbf{n}-\mathbf{n}_q+\mathbf{1}, \mathbf{n}-\mathbf{n}_q+\mathbf{3}, \cdots, \mathbf{n}-\mathbf{2}, \mathbf{n}, \\ \mathbf{n}_p+\mathbf{3}, \mathbf{n}_p+\mathbf{7}, \cdots, \mathbf{n}-\mathbf{n}_q-\mathbf{3}). \end{gathered}$$

(3) *If* $n-n_p-n_q\equiv 0 \pmod 4$, *then* $[\theta_p, s\theta_q^{-1}s] =$

$$(1,\mathbf{2})\cdot$$
$$(n_p-n_q-1, n-n_q-2, n-n_q-6,\cdots, n_p+2,$$
$$\mathbf{n}_p+\mathbf{5}, \mathbf{n}_p+\mathbf{9},\cdots, \mathbf{n}-\mathbf{n}_q-\mathbf{3},$$
$$n-1, n-3,\cdots, n-n_q+2, \mathbf{n}-\mathbf{n}_q+\mathbf{1}, \mathbf{n}-\mathbf{n}_q+\mathbf{3},\cdots, \mathbf{n}-\mathbf{2}, \mathbf{n},$$
$$\mathbf{n}_p+\mathbf{3}, \mathbf{n}_p+\mathbf{7},\cdots, \mathbf{n}-\mathbf{n}_q-\mathbf{5}, \mathbf{n}-\mathbf{n}_q-\mathbf{1},$$
$$n-n_q, n-n_q-4,\cdots, n_p+4, n_p, \mathbf{n}_p-\mathbf{n}_q+\mathbf{2})\cdot$$
$$(n_p-2, n_p-4,\cdots, n_p-n_q+1, \mathbf{n}_p-\mathbf{n}_q+\mathbf{4},$$
$$\mathbf{n}_p-\mathbf{n}_q+\mathbf{6},\cdots, \mathbf{n}_p-\mathbf{1}, \mathbf{n}_p+\mathbf{1}).$$

Proof. Note, $n_p-n_q>2$ implies $n_p>5$. By 10.4.5,

$$[\theta_p, s\theta_q^{-1}s] = g^{(n_q-n_p)}\theta_q^{-1}s\theta_q s g^{(n_p-n_q)}\cdot\theta_q s\theta_q^{-1}s$$

so by 10.4, $[\theta_p, s\theta_q^{-1}s] =$

$$(n_p, \mathbf{n}_p-\mathbf{n}_q, \mathbf{n}_p-\mathbf{n}_q-\mathbf{2},\cdots, \mathbf{2}, n-1, n-3,\cdots, n-n_q,\cdots, n_p+2,$$
$$\mathbf{n}_p+\mathbf{3}, \mathbf{n}_p+\mathbf{5},\cdots, \mathbf{n}-\mathbf{2}, 1, 3,\cdots, n_p-n_q-1, \mathbf{n}, \mathbf{n}_p+\mathbf{1})\cdot$$
$$(n_p-n_q+1, \mathbf{n}_p-\mathbf{n}_q+\mathbf{2})\cdot$$
$$(\mathbf{2}, \mathbf{4},\cdots, \mathbf{n}-\mathbf{n}_q-\mathbf{1}, n-1, 1, \mathbf{n}, n-n_q-2, n-n_q-4,\cdots, 3)\cdot$$
$$(n-n_q, \mathbf{n}-\mathbf{n}_q+\mathbf{1}).$$

Now we leave it for the reader to verify that the fixed points are as claimed.

Case 1. $\mathbf{n}-\mathbf{n}_q-\mathbf{n}_p-\mathbf{2}\equiv 0 \pmod 4$.

We write the cycles of $[\theta_p, s\theta_q^{-1}s]$ and let the reader verify the product. $[\theta_p, s\theta_q^{-1}s] =$

$$(1,\mathbf{2})\cdot$$
$$(n_p-n_q-1, n-n_q-2, n-n_q-6,\cdots, n_p, \mathbf{n}_p-\mathbf{n}_q+\mathbf{2})\cdot$$
$$(n_p-2, n_p-4,\cdots, n_p-n_q+1, \mathbf{n}_p-\mathbf{n}_q+\mathbf{4},$$
$$\mathbf{n}_p-\mathbf{n}_q+\mathbf{6},\cdots, \mathbf{n}_p-\mathbf{1}, \mathbf{n}_p+\mathbf{1})\cdot$$
$$(n-n_q, n-n_q-4,\cdots, n_p+6, n_p+2, \mathbf{n}_p+\mathbf{5}, \mathbf{n}_p+\mathbf{9},\cdots, \mathbf{n}-\mathbf{n}_q-\mathbf{1})\cdot$$
$$(n-1, n-3,\cdots, n-n_q+2, \mathbf{n}-\mathbf{n}_q+\mathbf{1}, \mathbf{n}-\mathbf{n}_q+\mathbf{3},\cdots, \mathbf{n}-\mathbf{2}, \mathbf{n},$$
$$\mathbf{n}_p+\mathbf{3}, \mathbf{n}_p+\mathbf{7},\cdots, \mathbf{n}-\mathbf{n}_q-\mathbf{3}).$$

Case 2. $\mathbf{n} - \mathbf{n}_p - \mathbf{n}_q \equiv 0 \pmod 4$.
$[\theta_p, s\theta_q^{-1}s] =$

$$(1, \mathbf{2}) \cdot$$
$$(n_p - n_q - 1, n - n_q - 2, n - n_q - 6, \cdots, n_p + 2,$$
$$\mathbf{n}_p + \mathbf{5}, \mathbf{n}_p + \mathbf{9}, \cdots, \mathbf{n} - \mathbf{n}_q - \mathbf{3},$$
$$n - 1, n - 3, \cdots, n - n_q + 2, \mathbf{n} - \mathbf{n}_q + \mathbf{1}, \mathbf{n} - \mathbf{n}_q + \mathbf{3}, \cdots, \mathbf{n} - \mathbf{2}, \mathbf{n},$$
$$\mathbf{n}_p + \mathbf{3}, \mathbf{n}_p + \mathbf{7}, \cdots, \mathbf{n} - \mathbf{n}_q - \mathbf{5}, \mathbf{n} - \mathbf{n}_q - \mathbf{1}, n - n_q,$$
$$n - n_q - 4, \cdots, n_p + 4, n_p, \mathbf{n}_p - \mathbf{n}_q + \mathbf{2}) \cdot$$
$$(n_p - 2, n_p - 4, \cdots, n_p - n_q + 1, \mathbf{n}_p - \mathbf{n}_q + \mathbf{4},$$
$$\mathbf{n}_p - \mathbf{n}_q + \mathbf{6}, \cdots, \mathbf{n}_p - \mathbf{1}, \mathbf{n}_p + \mathbf{1}).$$

10.6. *Suppose* $n_p - n_q = 2$. *Then:*

(1) *If* $n - 2n_p \equiv 2 \pmod 4$, *then* $[\theta_p, s\theta_q^{-1}s] =$

$$(1, n - n_p, n - n_p - 4, \cdots, n_p + 2, \mathbf{n}_p + \mathbf{5}, \mathbf{n}_p + \mathbf{9}, \cdots, \mathbf{n} - \mathbf{n}_p - \mathbf{1},$$
$$n - 1, n - 3, \cdots, n - n_p + 4, \mathbf{n} - \mathbf{n}_p + \mathbf{3}, \mathbf{n} - \mathbf{n}_p + \mathbf{5}, \cdots, \mathbf{n} - \mathbf{2}, \mathbf{n},$$
$$\mathbf{n}_p + \mathbf{3}, \mathbf{n}_p + \mathbf{7}, \cdots, \mathbf{n} - \mathbf{n}_p + \mathbf{1}, n - n_p + 2, n - n_p - 2, \cdots, n_p, \mathbf{4}, \mathbf{2}) \cdot$$
$$(\mathbf{6}, \mathbf{8}, \cdots, \mathbf{n}_p + \mathbf{1}, n_p - 2, , n_p - 4, \cdots, 5, 3).$$

(2) *If* $n - 2n_p \equiv 0 \pmod 4$, *then* $[\theta_p, s\theta_q^{-1}s] =$

$$(1, n - n_p, n - n_p - 4, \cdots, n_p, \mathbf{4}, \mathbf{2}) \cdot$$
$$(\mathbf{6}, \mathbf{8}, \cdots, \mathbf{n}_p + \mathbf{1}, n_p - 2, n_p - 4, \cdots, 5, 3) \cdot$$
$$(n - 1, n - 3, \cdots, n - n_p + 4, \mathbf{n} - \mathbf{n}_p + \mathbf{3}, \mathbf{n} - \mathbf{n}_p + \mathbf{5},$$
$$\cdots, \mathbf{n} - \mathbf{2}, \mathbf{n}, \mathbf{n}_p + \mathbf{3}, \mathbf{n}_p + \mathbf{7}, \cdots, \mathbf{n} - \mathbf{n}_p - \mathbf{1}) \cdot$$
$$(n - n_p + 2, n - n_p - 2, \cdots, n_p + 2, \mathbf{n}_p + \mathbf{5}, \mathbf{n}_p + \mathbf{9}, \cdots, \mathbf{n} - \mathbf{n}_p + \mathbf{1}).$$

Proof. By 10.4.5, $[\theta_p, s\theta_q^{-1}s] =$

$$g^{(n_q - n_p)} \theta_q^{-1} s \theta_q s g^{(n_p - n_q)} \cdot$$
$$\theta_q s \theta_q^{-1} s$$

so by 10.4, (replacing n_q by $n_p - 2$), $[\theta_p, s\theta_q^{-1}s] =$

$$(\mathbf{n}_p, \mathbf{2}, n-1, n-3, \cdots, n-n_p+2, \cdots, n_p+2,$$
$$\mathbf{n}_p+\mathbf{3}, \mathbf{n}_p+\mathbf{5}, \cdots, \mathbf{n}-\mathbf{2}, 1, \mathbf{n}, \mathbf{n}_p+\mathbf{1})(3, \mathbf{4})\cdot$$
$$(\mathbf{2}, \mathbf{4}, \cdots, \mathbf{n}-\mathbf{n}_p+\mathbf{1}, n-1, 1, \mathbf{n}, n-n_p, n-n_p-2, \cdots, 3)\cdot$$
$$(n-n_p+2, \mathbf{n}-\mathbf{n}_p+\mathbf{3}).$$

Case 1. $\mathbf{n}-\mathbf{2n}_p-\mathbf{2} \equiv 0 \pmod 4$.

We write the cycles of $[\theta_p, s\theta_q^{-1}s]$ and let the reader verify the product.

$$[\theta_p, s\theta_q^{-1}s] =$$
$$(1, n-n_p, n-n_p-4, \cdots, n_p+2, \mathbf{n}_p+\mathbf{5}, \mathbf{n}_p+\mathbf{9}, \cdots, \mathbf{n}-\mathbf{n}_p-\mathbf{1},$$
$$n-1, n-3, \cdots, n-n_p+4, \mathbf{n}-\mathbf{n}_p+\mathbf{3}, \mathbf{n}-\mathbf{n}_p+\mathbf{5}, \cdots, \mathbf{n}-\mathbf{2}, \mathbf{n},$$
$$\mathbf{n}_p+\mathbf{3}, \mathbf{n}_p+\mathbf{7}, \cdots, \mathbf{n}-\mathbf{n}_p+\mathbf{1}, n-n_p+2, n-n_p-2, \cdots, n_p, \mathbf{4}, \mathbf{2})\cdot$$
$$(\mathbf{6}, \mathbf{8}, \cdots, \mathbf{n}_p+\mathbf{1}, n_p-2, n_p-4, \cdots, 5, 3).$$

Case 2. $\mathbf{n}-\mathbf{2n}_p \equiv 0 \pmod 4$

$$[\theta_p, s\theta_q^{-1}s] =$$
$$(1, n-n_p, n-n_p-4, \cdots, n_p, \mathbf{4}, \mathbf{2})\cdot$$
$$(\mathbf{6}, \mathbf{8}, \cdots, \mathbf{n}_p+\mathbf{1}, n_p-2, n_p-4, \cdots, 5, 3)\cdot$$
$$(n-1, n-3, \cdots, n-n_p+4, \mathbf{n}-\mathbf{n}_p+\mathbf{3}, \mathbf{n}-\mathbf{n}_p+\mathbf{5},$$
$$\cdots, \mathbf{n}-\mathbf{2}, \mathbf{n}, \mathbf{n}_p+\mathbf{3}, \mathbf{n}_p+\mathbf{7}, \cdots, \mathbf{n}-\mathbf{n}_p-\mathbf{1})\cdot$$
$$(n-n_p+2, n-n_p-2, \cdots, n_p+2, \mathbf{n}_p+\mathbf{5}, \mathbf{n}_p+\mathbf{9}, \cdots, \mathbf{n}-\mathbf{n}_p+\mathbf{1}).$$

We can now complete the proof of Theorem 10.2.

Proof of Theorem 10.2. First we show that (1) implies (2). Since Γ is transitive, $C_G(\Gamma)$ is a semi-regular subgroup of G. But $[\theta_p, C_G(\Gamma)] = 1$, and θ_p has a single fixed point, hence $C_G(\Gamma) = 1$.

We proceed with the proof of (1). Assume first that $p = q$. Then $\theta_q s\theta_q^{-1}s \in \Gamma$. Recall from 10.4 that

$$\theta_q s\theta_q^{-1}s =$$
$$(\mathbf{2}, \mathbf{4}, \cdots, \mathbf{n}-\mathbf{n}_q-\mathbf{1}, n-1, 1, \mathbf{n}, n-n_q-2, n-n_q-4,$$
$$\cdots, 3)(n-n_q, \mathbf{n}-\mathbf{n}_q+\mathbf{1}).$$

Hence $\{1, \mathbf{2}, 3, \mathbf{4}, \cdots, \mathbf{n}-\mathbf{n}_q-\mathbf{1}\}$ are in the same orbit of Γ. However, since $q \geq 3$, $n-n_q-1 > n_q$, and the above set contains a representative from each orbit of θ_q. Hence $\{1, 2, \cdots, n-1\}$ are in the same orbit of Γ, and looking at $\theta_q s\theta_q^{-1}s$, we see that n is also there.

Suppose next that $n_p - n_q > 2$. Note that $[\theta_p, s\theta_q^{-1}s] \in \Gamma$. Assume first that $n - n_q - n_p \equiv 2 \pmod 4$. We use 10.5.2. We write the cycles in $[\theta_p, s\theta_q^{-1}s]$

$$\sigma_1 = (1, \mathbf{2})$$

$$\sigma_2 = (n_p - n_q - 1, n - n_q - 2, n - n_q - 6, \cdots, n_p, \mathbf{n}_p - \mathbf{n}_q + \mathbf{2})$$

$$\sigma_3 = (n_p - 2, n_p - 4, \cdots, n_p - n_q + 1,$$
$$\mathbf{n}_p - \mathbf{n}_q + \mathbf{4}, \mathbf{n}_p - \mathbf{n}_q + \mathbf{6}, \cdots, \mathbf{n}_p - \mathbf{1}, \mathbf{n}_p + \mathbf{1})$$

$$\sigma_4 = (n - n_q, n - n_q - 4, \cdots, n_p + 4, n_p + 2,$$
$$\mathbf{n}_p + \mathbf{5}, \mathbf{n}_p + \mathbf{9}, \cdots, \mathbf{n} - \mathbf{n}_q + \mathbf{1})$$

$$\sigma_5 = (n - 1, n - 3, \cdots, n - n_q + 2, \mathbf{n} - \mathbf{n}_q + \mathbf{1}, \mathbf{n} - \mathbf{n}_q + \mathbf{3},$$
$$\cdots, \mathbf{n} - \mathbf{2}, \mathbf{n}, \mathbf{n}_p + \mathbf{3}, \mathbf{n}_p + \mathbf{7}, \cdots, \mathbf{n} - \mathbf{n}_q - \mathbf{3}).$$

Recall that the orbits of θ_p are

$$X_i = \{i, n_p + i, 2n_p + i, \cdots, (p-1)n_p + i\}, \quad 1 \le i \le n_p.$$

Let $\mathcal{O}$ be the orbit of 1 (under Γ), then $\operatorname{supp}(\sigma_1) \subseteq \mathcal{O}$. Note that $1, n_p + 1 \in X_1$ hence $\operatorname{supp}(\sigma_3) \subseteq \mathcal{O}$. Note that $n_p - 1, n - 2 \in X_{n_p - 1}$, hence $\operatorname{supp}(\sigma_5) \subseteq \mathcal{O}$. Note that $2, n_p + 2 \in X_2$, hence $\operatorname{supp}(\sigma_4) \subseteq \mathcal{O}$. Also $n_p, n - 1 \in X_{n_p}$, hence $\operatorname{supp}(\sigma_2) \subseteq \mathcal{O}$. Since no two elements in $\operatorname{Fix}([\theta_p, s\theta_q^{-1}s])$, are in the same orbit of θ_p, $\mathcal{O} = \{1, 2, \cdots, n\}$ and Γ is transitive.

Assume next that $n - n_q - n_p \equiv 0 \pmod 4$. We use 10.5.3. We write the cycles in $[\theta_p, s\theta_q^{-1}s]$

$$\gamma_1 = (1, \mathbf{2}).$$

$$\gamma_2 = (n_p - n_q - 1, n - n_q - 2, n - n_q - 6, \cdots, n_p + 2,$$
$$\mathbf{n}_p + \mathbf{5}, \mathbf{n}_p + \mathbf{9}, \cdots, \mathbf{n} - \mathbf{n}_q - \mathbf{3},$$
$$n - 1, n - 3, \cdots, n - n_q + 2, \mathbf{n} - \mathbf{n}_q + \mathbf{1}, \mathbf{n} - \mathbf{n}_q + \mathbf{3}, \cdots, \mathbf{n} - \mathbf{2}, \mathbf{n},$$
$$\mathbf{n}_p + \mathbf{3}, \mathbf{n}_p + \mathbf{7}, \cdots, \mathbf{n} - \mathbf{n}_q - \mathbf{5}, \mathbf{n} - \mathbf{n}_q - \mathbf{1},$$
$$n - n_q, n - n_q - 4, \cdots, n_p + 4, n_p, \mathbf{n}_p - \mathbf{n}_q + \mathbf{2}).$$

$$\gamma_3 = (n_p - 2, n_p - 4, \cdots, n_p - n_q + 1,$$
$$\mathbf{n}_p - \mathbf{n}_q + \mathbf{4}, \mathbf{n}_p - \mathbf{n}_q + \mathbf{6}, \cdots, \mathbf{n}_p - \mathbf{1}, \mathbf{n}_p + \mathbf{1}).$$

Let $\mathcal{O}$ be the orbit of 1. Then $\operatorname{supp}(\gamma_1) \subseteq \mathcal{O}$. Then, as $1,\ n_p+1 \in X_1$, $\operatorname{supp}(\gamma_3) \subseteq \mathcal{O}$, and as $2,\ n_p+2 \in X_2$, $\operatorname{supp}(\gamma_2) \subseteq \mathcal{O}$, so as above, $\mathcal{O} = \{1, 2, \cdots, n\}$.

Finally, suppose that $n_p - n_q = 2$. Assume first that $n - 2n_p \equiv 2 \pmod 4$. We use 10.6.1. We write the cycles in $[\theta_p, s\theta_q^{-1}s]$

$$\alpha_1 = (1, n-n_p, n-n_p-4, \cdots, n_p+2, \mathbf{n}_p+\mathbf{5}, \mathbf{n}_p+\mathbf{9}, \cdots, \mathbf{n}-\mathbf{n}_p-\mathbf{1},$$
$$n-1, n-3, \cdots, n-n_p+4, \mathbf{n}-\mathbf{n}_p+\mathbf{3}, \mathbf{n}-\mathbf{n}_p+\mathbf{5}, \cdots, \mathbf{n}-\mathbf{2}, \mathbf{n},$$
$$\mathbf{n}_p+\mathbf{3}, \mathbf{n}_p+\mathbf{7}, \cdots, \mathbf{n}-\mathbf{n}_p+\mathbf{1}, n-n_p+2, n-n_p-2, \cdots, n_p, \mathbf{4}, \mathbf{2}).$$
$$\alpha_2 = (\mathbf{6}, \mathbf{8}, \cdots, \mathbf{n}_p+\mathbf{1}, n_p-2, n_p-4, \cdots, 5, 3).$$

Let $\mathcal{O}$ be the orbit of 1. Then $\operatorname{supp}(\alpha_1) \subseteq \mathcal{O}$. Then as $1,\ n_p+1 \in X_1$, $\operatorname{supp}(\alpha_2) \subseteq \mathcal{O}$ so $\mathcal{O} = \{1, 2, \cdots, n\}$.

Finally, assume that $n - 2n_p \equiv 0 \pmod 4$. We use 10.6.2. We write the cycles in $[\theta_p, s\theta_q^{-1}s]$

$$\beta_1 = (1, n-n_p, n-n_p-4, \cdots, n_p, \mathbf{4}, \mathbf{2})$$
$$\beta_2 = (\mathbf{6}, \mathbf{8}, \cdots, \mathbf{n}_p+\mathbf{1}, n_p-2, n_p-4, \cdots, 5, 3)$$
$$\beta_3 = (n-1, n-3, \cdots, n-n_p+4,$$
$$\mathbf{n}-\mathbf{n}_p+\mathbf{3}, \mathbf{n}-\mathbf{n}_p+\mathbf{5}, \cdots, \mathbf{n}-\mathbf{2}, \mathbf{n}, \mathbf{n}_p+\mathbf{3}, \mathbf{n}_p+\mathbf{7}, \cdots, \mathbf{n}-\mathbf{n}_p-\mathbf{1})$$
$$\beta_4 = (n-n_p+2, n-n_p-2, \cdots, n_p+2, \mathbf{n}_p+\mathbf{5}, \mathbf{n}_p+\mathbf{9}, \cdots, \mathbf{n}-\mathbf{n}_p+\mathbf{1}).$$

Let $\mathcal{O}$ be the orbit of 1. Then $\operatorname{supp}(\beta_1) \subseteq \mathcal{O}$. Then as $1,\ n_p+1 \in X_1$, $\operatorname{supp}(\beta_2) \subseteq \mathcal{O}$, and as $3,\ n_p+3 \in X_3$, $\operatorname{supp}(\beta_3) \subseteq \mathcal{O}$. Now, since 2, $n_p+2 \in X_2$, $\operatorname{supp}(\beta_4) \subseteq \mathcal{O}$, so $\mathcal{O} = \{1, 2, \cdots, n\}$. This completes the proof of Theorem 10.2.

11. The Sporadic Groups.

In this short section we point out the following theorem.

Theorem 11.1. *Let L be a Sporadic finite simple group. Then $\Delta(L)$ is disconnected.*

Proof. Let L be a sporadic group. We show that there exists a prime $p = p(L)$, such that if $x \in L$ is an element of order p, then $C_L(x) = \langle x \rangle$. Of course $\langle x \rangle - \{1\}$ is a connected component of $\Delta(L)$. We use the Atlas [**2**]. The following table gives the value of $p(L)$.

L	$p(L)$	L	$p(L)$	L	$p(L)$
M_{11}	11	M_{12}	11	M_{22}	11
M_{23}	23	M_{24}	23	Co_1	23
Co_2	23	Co_3	23	J_1	19
J_2	7	J_3	19	J_4	43
Fi_{22}	13	Fi_{23}	23	Fi'_{24}	29
F_1	71	F_2	47	F_3	31
F_5	19	He	17	McL	11
HS	11	Suz	13	O'N	31
Ly	67	Ru	29		

12. Concluding results.

In this section we prove Theorem 4 of the introduction and present related results on division algebras. In addition, we include a number of results and remarks related to the commuting graph of the classical groups. Throughout $\mathfrak{G}$ will denote a connected reductive algebraic group over an algebraically closed field defined over an infinite field K. Let $\mathfrak{G}(K)$ denote the K rational points.

12.1. ([**10**, Thm. 2.2].) Let $\mathfrak{G}$ be a connected nonabelian reductive group defined over an infinite field K. Then $\mathfrak{G}(K)$ is Zariski dense in $\mathfrak{G}$.

12.2. *Let K be an abelian field and $\mathfrak{G}$ a nonabelian reductive algebraic group defined over K. Then:*

(1) $\mathfrak{G}(K)/Z(\mathfrak{G}(K))$ *does not have finite exponent.*
(2) *Let $Z \leq Z(\mathfrak{G}(K))$. If A/Z is an abelian normal subgroup of $\mathfrak{G}(K)/Z$, then $A \leq Z(\mathfrak{G}(K))$.*
(3) $\mathfrak{G}(K)$ *is not solvable.*

Proof. By 12.1, $\mathfrak{G}(K)$ is Zariski dense in $\mathfrak{G}$. As centralizers of elements in $\mathfrak{G}$ are Zariski closed, it follows that $Z(\mathfrak{G}(K)) \leq Z(\mathfrak{G})$. Then $\mathfrak{G}(K)/Z(\mathfrak{G}(K))$ is Zariski dense in $\mathfrak{G}/Z(\mathfrak{G}(K))$.

(1): If $\mathfrak{G}/Z(\mathfrak{G}(K))$ has exponent n, then, as the set of elements of order n in $\mathfrak{G}/Z(\mathfrak{G}(K))$ is Zariski closed, this forces $\mathfrak{G}/Z(\mathfrak{G}(K))$ to be of finite exponent. But this is clearly false as seen by considering a torus.

Let $Z \leq Z(\mathfrak{G}(K))$ and suppose $1 < A/Z \lhd \mathfrak{G}(K)/Z$ with A/Z abelian. The Zariski closure, say B/Z, of A/Z in $\mathfrak{G}/Z$ is abelian (indeed the center of $\bigcap_{a\in A} C_{\mathfrak{G}/Z}(Za)$ is a closed abelian subgroup of $\mathfrak{G}/Z$ containing A/Z). Also B/Z is normalized by $\mathfrak{G}(K)/Z$. Now normalizers are closed, so B/Z is an abelian normal closed subgroup in $\mathfrak{G}/Z$. But as $\mathfrak{G}$ is a connected reductive group, $B \leq Z(\mathfrak{G})$, a contradiction. This proves (2) and (3) follows.

Corollary 12.3. *Let D be a division algebra over K. Then D^* is not solvable.*

Proof. This follows from 12.2.3 by noting that D^* can be realized as the K rational points of GL_d, where $d = \deg(D)$.

We can now derive Theorem 4 of the introduction.

Theorem 12.4. *Let D be a finite dimensional division algebra over a number field K. Let N be a noncentral normal subgroup of D^*. Then D^*/N solvable.*

Proof. Let $S := SL_1(D)$ be the elements of D^* whose reduced norm is 1. Then $N/(N \cap S) \cong NS/S$ is abelian, so by 12.2.2, $N \cap S$ is noncentral in D^* (alternatively, use [**13**]).

Hence it suffices to show that if M is a noncentral normal subgroup of $SL_1(D)$, then $SL_1(D)/M$ is solvable. Here we take $\mathfrak{G}$ a simple, simply connected algebraic group of type A_n such that $\mathfrak{G}(K) = SL_1(D)$.

Suppose $M \lhd \mathfrak{G}(K)$ and M is not central. We apply Theorem 2 (of the introduction). If $T = \emptyset$, then $M = \mathfrak{G}(K)$ and there is nothing to prove. Thus we suppose $T \neq \emptyset$. Hence we can consider $\mathfrak{G}(K) < \prod_{v \in T} \mathfrak{G}(K_v)$, via the diagonal embedding. By Theorem 2, $M = \mathfrak{G}(K) \cap L$, where $L \lhd \prod_{v \in T} \mathfrak{G}(K_v)$, with L open. Then $\mathfrak{G}(K)/M = \mathfrak{G}(K)/(\mathfrak{G}(K) \cap L) \cong \mathfrak{G}(K)L/L$ and so it suffices to show that $\prod_{v \in T} \mathfrak{G}(K_v)/L$ is solvable.

Notice that for each $v \in T$, $[\mathfrak{G}(K_v), L] \leq \mathfrak{G}(K_v) \cap L$ is a normal subgroup of $\mathfrak{G}(K_v)$ and of course $\prod_{v \in T} \mathfrak{G}(K_v)/L$ is an image of $\prod_{v \in T}(\mathfrak{G}(K_v)/[\mathfrak{G}(K_v), L])$. So it suffices to show that $\mathfrak{G}(K_v)/[\mathfrak{G}(K_v), L]$ is solvable. Let M_v (resp. L_v) be the projection of M (resp. L) on $\mathfrak{G}(K_v)$. Since M is noncentral in $\mathfrak{G}(K)$, M_v and hence L_v is noncentral in $\mathfrak{G}(K_v)$. Then, by 12.2.2, $[\mathfrak{G}(K_v), L] = [\mathfrak{G}(K_v), L_v]$ is noncentral in $\mathfrak{G}(K_v)$. Then, by [**12**] (see also [**10**, Prop. 1.8, p. 32]), $[\mathfrak{G}(K_v), L]$ contains C_s, for some s, where C_s are the congruence subgroups of $\mathfrak{G}(K_v) = SL_1(D_v)$ (where $D_v = D \otimes_K K_v$). These congruence subgroups are defined in [**10**, p. 31 (1.4.4)]. Since $\mathfrak{G}(K_v)/C_s$ is solvable ([**10**, Corollary, p. 32]), we are done.

Next we focus our attention on the commuting graph of the classical groups. We mention that as noted in Theorem 5 of the Introduction, the elements x, y required for showing that $\Delta(L)$ is balanced can be taken as opposite unipotent elements. We remark that except for some small cases this usually implies $d(x, y) = 4$. To see this note that $C_L(x), C_L(y)$ contain root elements r, s lying in root groups corresponding to opposite long roots of the root system. The normalizer of these root groups are opposite parabolic subgroups, hence contain a common Levi factor. Choosing $1 \neq t$ in this Levi factor (which is possible in all but a few cases) we have a path x, r, t, s, y of length 4.

In the following theorem we use the same ϵ notation as given in the beginning of Section 9.

Theorem 12.5. *Let $G(q)$ be a simple classical group with $q > 5$. Then $\Delta(G(q))$ is disconnected if and only if one of the following holds:*

(i) $G(q) \simeq L_n^\epsilon(q)$ *and n is a prime.*
(ii) $G(q) \simeq L_n^\epsilon(q)$, *$n-1$ is a prime and $q - \epsilon \mid n$.*
(iii) $G(q) \simeq S_{2n}(q)$, $O_{2n}^-(q)$, *or* $O_{2n+1}(q)$ *and $n = 2^c$, for some c.*

Moreover, if $\Delta(G(q))$ is connected then $\operatorname{diam}(\Delta(G(q))) \leq 10$.

Proof. Let $\hat{G}(q)$ denote the corresponding quasisimple classical group and let V be the natural module for $\hat{G}(q)$. For a nondegenerate subspace $W \leq V$, we write $I(W)$ for $GL(W), GU(W), Sp(W)$ or $SO(W)$, in the respective cases. We let $\hat{G}(W) \leq \hat{G}(q)$ be the subgroup acting trivially on $W^\perp$ (and acting trivially on a specified complement U, in the case when $\hat{G}(q) \simeq SL_n(q)$, the complement U in this case will be clear from the context).

For the orthogonal groups we assume that $\dim(V) \geq 7$. First suppose that $G(q)$ does not satisfy any of the conditions (i)-(iii). Here we will show that $\operatorname{diam}(\Delta(G(q))) \leq 10$. The following is the key step.

($*$) Each $g \in G(q)$ is at distance at most 3 from some unipotent element in $\Delta(G(q))$.

We proceed by contradiction assuming that ($*$) does not hold. If g is the commuting product of a nontrivial unipotent element and a semisimple element, then ($*$) is obvious. Therefore g is a semisimple element.

Let h be a preimage of g in $\hat{G}(q)$. Then h is contained in a maximal torus T of $I(V)$. When $I(V) \simeq SO_{2n+1}(q)$, all maximal tori are contained in $SO_{2n}^\epsilon(q)$, for $\epsilon = 1$ or -1, so here all considerations can be reduced to even dimensional orthogonal groups and we therefore ignore odd dimensional orthogonal groups in the following.

The action of T on V is completely reducible and given by Lemma 2 of [**16**] (the $q > 5$ hypothesis is sufficient to establish that lemma). Alternatively, one can obtain a suitable torus working directly from a decomposition of V under the action of h. In any case, T preserves a decomposition $V = V_1 \perp \ldots \perp V_k \perp (V_{k+1} \oplus V'_{k+1}) \perp \ldots \perp (V_\ell \oplus V'_\ell)$, where if we set $\dim(V_i) = r_i$, $1 \leq i \leq \ell$, then $r_1 \geq \ldots \geq r_k$, and for $k < i \leq \ell$, $\dim(V_i) = \dim(V'_i)$, with both subspaces being totally singular.

Corresponding to this decomposition we have $T = T_1 \times \cdots \times T_\ell$, such that for $1 \leq i \leq \ell$, T_i induces a Singer cycle on V_i and for $k < i \leq \ell$, T_i also induces a Singer cycle on V'_i. We note that $k = \ell$ in the general linear case. Also for $1 \leq i \leq k$, one of the following holds: $|T_i| = q^{r_i} - 1$, $q^{r_i} + 1$ (with r_i odd), $q^{r_i/2} + 1$, $q^{r_i/2} + 1$, with $I(V_i) = GL_{r_i}(q), GU_{r_i}(q), Sp_{r_i}(q)$, or $SO^-_{r_i}(q)$, respectively. We make a series of reductions under the assumption that $(*)$ fails to hold for g.

Step 1. $\dim(V_i) = 1$, for each $i > k$.

For suppose $k < i \leq \ell$ and $\dim(V_i) > 1$, $T_i \leq GL_{r_i}(q)$ ($GL_{r_i}(q^2)$ in the unitary case) with dual action on V_i and V_i'. Then T_i contains a subgroup Z_i of order $q-1$ (q^2-1 in the unitary case) which induces (inverse) scalars on V_i, and V_i'. Elements of Z_i have determinant 1 and since we are assuming $q > 5$, we can find a noncentral element of Z_i in $\hat{G}(q)$. Since all elements of this group centralize unipotent elements of $GL_{r_i}(q)$, we obtain $(*)$ in this case, a contradiction.

Step 2. $\ell \leq k+1$, if $G(q) \neq O^{\epsilon}_{2n}(q)$. Otherwise $\ell \leq k+2$.

For suppose $\ell > k$. Then Z_{k+1} centralizes $\hat{G}(V_{k+2} \oplus \cdots \oplus V_\ell')$, so this group contains no unipotent elements. Hence either $\ell = k+1$, or $G(q)$ is an orthogonal group and $\ell = k+2$.

Step 3. $k = \ell$.

First assume $k = 0$. Then Step 1 and Step 2 show that either $\dim(V) = 2$, or $\dim(V) = 4$, with $G(q) \simeq O_4^-(q)$ (as $G(q)$ is simple). In either case (i) or (iii) holds, a contradiction. Now suppose $0 < k < \ell$. Then Z_ℓ commutes with $\hat{G}(V_1 \oplus \cdots \oplus V_k)$ and the latter group contains unipotent elements unless either $V_1 \oplus \cdots \oplus V_k$ is a 2-dimensional orthogonal space or a 1-dimensional unitary space (we already mentioned that $k = \ell$ if $G(q) \simeq L_n(q)$). In the former case Step 2 implies $\dim(V) \leq 6$, against our supposition. And in the unitary case, $\dim(V) = 3$ and hence satisfies (i). This is again a contradiction.

Step 4. $r_1 > 1$.

Suppose $r_1 = 1$. This can only occur for $G(q) = L^{\epsilon}_n(q)$. We are assuming that (i) does not hold, so here $k = n \geq 4$. Then $(T_1 \times T_2) \cap \hat{G}(q)$ contains a noncentral subgroup of order $q - \epsilon$ centralizing unipotent elements in $\hat{G}(V_3 \oplus \cdots \oplus V_k)$, a contradiction.

Step 5. Either $V = V_1$ or $G(q) = L^{\epsilon}_n(q)$, $V = V_1 \oplus V_2$, and $\dim(V_2) = 1$.

It follows from Step 4 that T_1 contains noncentral elements of $\hat{G}(q)$. Since we are assuming that $(*)$ does not hold, $\hat{G}(V_2 \oplus \cdots \oplus V_k)$ contains no non-identity unipotent elements.

If $G = L^{\epsilon}_n(q)$, this forces $\dim(V_2 \oplus \cdots \oplus V_k) \leq 1$. In the symplectic case, necessarily $V = V_1$. We argue that this holds for the orthogonal case as well. For otherwise, $k = 2$ and $\dim(V_2) = 2$. Hence $\dim(V_1) \geq 5$. But then there are noncentral elements of T_2 which centralize unipotent elements of $\hat{G}(V_1)$, a contradiction.

We now treat the remaining configurations. First assume $V = V_1$, so that $r_1 = n$. If $G(q) = L^{\epsilon}_n(q)$, then $|T| = q^n - \epsilon$. Also n is odd in the unitary case.

We are assuming that n is not a prime, so we may write $n = rs$, with $r, s > 1$ and such that s is odd in the unitary case. Then there is a (cyclic) subgroup $E < T$ of order $q^r - \epsilon$ intersecting $\hat{G}(q)$ in a noncentral subgroup. As T acts irreducibly on V, E acts homogeneously, so that $V = W_1 \oplus \cdots \oplus W_s$, with each W_i of dimension r and irreducible under the action of E. In the unitary case where s is odd, it is easily checked that we may take W_1 nondegenerate and perpendicular to the remaining summands. Now h centralizes E which in turn centralizes a Singer cycle in $\hat{G}(W_1)$. This Singer cycle centralizes a unipotent element in $\hat{G}(W_2 \oplus \cdots \oplus W_s)$ so we have $(*)$, a contradiction.

In the symplectic and orthogonal cases, we have $|T| = q^n + 1$. Here we are assuming that n is not a power of 2, so the same argument works.

The final case is where $V = V_1 \oplus V_2$, with $\dim(V_2) = 1$ and $G = L_n^\epsilon(q)$. Then $r_1 = n - 1$. If $n - 1$ is not a prime, we argue as above, working in $SL_{n-1}^\epsilon(q)$. Suppose $n - 1$ is a prime. Then T contains a subgroup of order $(q - \epsilon)^2$ which induces scalars on V_i. Intersecting with $\hat{G}(q)$ we get a group of order $q - \epsilon$ so this gives a noncentral element centralizing a unipotent element of $\hat{G}(V_1)$, unless $q - \epsilon \mid n$. This concludes the proof of $(*)$.

It is now an easy matter to show that $\Delta(G(q))$ is connected of diameter at most 10. By $(*)$ g is at distance at most 3 from a nontrivial unipotent element of $G(q)$. The center of a maximal unipotent subgroup of $G(q)$ contains long root elements. Hence g is at distance at most 4 from a long root element.

Now let $g, g' \in \Delta(G(q))$. Let u, u' be long root elements at distance at most 4 from g, g' respectively. It is well-known that either u, u' commute, lie in an extraspecial p-subgroup (hence commute with the center), or lie in a group $J = SL_2(q)$ generated by the long root subgroups corresponding to u, u'. In the latter case, we can choose a root element w lying in a conjugate of J and commuting with J. This completes the argument.

To complete the proof of the theorem we now assume that $G(q)$ satisfies either (i), (ii) or (iii). Here we argue that $\Delta(G(q))$ is disconnected. If (i) holds with $n = p$ a prime, then $GL_p^\epsilon(q)$ contains a cyclic maximal torus T of order $q^p - \epsilon$. If $p = 2$, then we immediately see that opposite unipotent elements cannot be joined. So assume p is odd. Let $h \in E = T \cap SL_p^\epsilon(q)$ with $h \notin Z(SL_p^\epsilon(q))$. So h acts irreducibly on V. Suppose $y \in SL_p^\epsilon(q)$ centralizes h projectively. Hence $h^y = hz$, where $z \in Z(SL_p^\epsilon(q))$. The centralizer of h and of h^y in $SL_p^\epsilon(q)$ is E, so y normalizes E, hence induces an automorphism on E of order dividing p. Hence z has order dividing $(p, q - \epsilon)$. So either $z = 1$, or is of order p. In the latter case, by 8.3, $|E/(E \cap Z(SL_p^\epsilon(q)))|$ has order prime to p, so we may assume h has order prime to p, and this also forces $z = 1$. But the centralizer of h in $SL_p^\epsilon(q)$ is E, so the image of $E - \{1\}$ in $G(q)$ is a connected component of $\Delta(G(q))$.

The same argument applies if (iii) holds, taking T to be a Singer cycle of order q^n+1 and noting that the resulting torus of the simple group has odd order.

The last case is where (ii) holds with $n-1=p$ a prime and $q-\epsilon$ dividing $n=p+1$. In this case take a decomposition $V=V_1 \perp V_2$, with $\dim(V_1)=p$. Then $GL_n^\epsilon(q)$ contains a maximal torus $T_1 \times T_2$ of order $(q^p-\epsilon)(q-\epsilon)$. The resulting torus $E < SL_n^\epsilon(q)$ has order $(q^p-\epsilon)$ and in the simple group the torus has order $(q^p-\epsilon)/(q-\epsilon)$. The argument is thus the same as in the case where (i) holds. This completes the proof of Theorem 12.5.

Remarks. (1) In the papers [**19**] and [**4**] the connected components of the prime graph of all nonabelian finite simple groups are determined. It is easy to see that the prime graph is connected if and only if the commuting graph is connected. Thus the nonabelian finite simple groups L for which $\Delta(L)$ is disconnected are known. We note that in the connected case of Theorem 12.2 we prove that the diameter of $\Delta(G(q))$ is bounded.

(2) We assume $q>5$, in the above result, in order to simplify the statement and the proof. With extra work one should be able to obtain information for smaller values of q. However, there will be additional examples where the graph is disconnected.

References

[1] A. Borel, R. Carter, N. Iwahori, T.A. Springer and R. Steinberg, *Seminar on Algebraic Groups and Related Finite Groups*, Lecture Notes in Math., **131**, Springer-Verlag, 1970, MR 48 #11106, Zbl 192.36201.

[2] J.H. Conway et al, *Atlas of Finite Groups*, Clarendon Press, Oxford, 1985, MR 88g:20025, Zbl 568.20001.

[3] R.H. Dye, *A geometric characterization of the special orthogonal groups and the Dickson invariant*, J. London Math. Soc., **15(2)** (1977), 472-476, MR 56 #11899, Zbl 358.20054.

[4] N. Iiyori and H. Yamaki, *Prime graph components of the simple groups of Lie type over the fields of even characteristic*, J. Algebra, **115** (1993), 335-343, MR 94e:05268, Zbl 799.20016; *Corrections*, J. Algebra, **181** (1996), 659, CMP 1 383 487.

[5] P. Kleidman, *The maximal subgroups of the Steinberg triality groups* $^3D_4(q)$ *and their automorphism groups*, J. Algebra, **115(1)** (1988), 182-199, MR 89f:20024, Zbl 642.20013.

[6] M. Liebeck and G. Seitz, *Reductive subgroups of exceptional algebraic groups*, Memoirs of the AMS, **121** (1996), MR 96i:20059, Zbl 851.20045.

[7] K. Mizuno, *The conjugate classes of Chevalley groups of type* E_6, J. Fac. Sci. Tokyo Univ. IA-Math., **24** (1977), 525-563, MR 58 #5951, Zbl 399.20044.

[8] ———, *The conjugate classes of unipotent elements of Chevalley groups* E_7 *and* E_8, Tokyo J. Math., **3(2)** (1980), 391-459, MR 82m:20046, Zbl 454.20046.

[9] K. Motose, *On values of cyclotomic polynomials* II, Math. J. of Okayama Univ., **37** (Dec. 1995), 27-36, MR 97h:11151.

[10] V. Platonov and A. Rapinchuk, *Algebraic Groups and Number Theory*, 'Nauka' Publishers, Moscow, 1991; English translation: Pure and Applied Mathematics series, **139**, Academic Press, 1994, MR 95b:11039, Zbl 732.20027.

[11] A. Potapchik and A. Rapinchuk, *Normal subgroups of $SL_{1,D}$ and the classification of finite simple groups*, Proc. Indian Acad. Sci. (Math. Sci.), **106**(**4**) (1996), 329-368, MR 98i:20049, Zbl 879.20027.

[12] C. Riehm, *The norm* 1 *group of p-adic division algebras*, Amer. J. Math., **92**(**2**) (1970), 499-523, MR 1 #6860, Zbl 199.37601.

[13] W.R. Scott, *On the multiplicative group of a division ring*, Proc. Amer. Math. Soc., **8** (1957), 303-305, MR 18,788g, Zbl 079.05402.

[14] Y. Segev, *On finite homomorphic images of the multiplicative group of a division algebra*, Ann. of Math., **149** (1999), 219-251, MR 2000e:16022, Zbl 935.20009.

[15] G.M. Seitz, *Generation of finite groups of Lie type*, Trans. AMS, **271**(**2**) (1982), 351-407, MR 83h:20021, Zbl 514.20013.

[16] ______, *On the subgroup structure of classical groups*, Communications in Algebra, **10** (1982), 875-885, MR 83k:20052, Zbl 483.20026.

[17] K. Shinoda, *The conjugacy classes of the finite Ree groups of type* (F_4), J. Fac. Sci. Tokyo Univ. IA-Math., **22** (1975), 1-15, MR 51 #8281, Zbl 306.20014.

[18] P.H. Tiep and A.E. Zalesskii, *Minimal characters of the finite classical groups*, Comm. in Alg., **24**(**6**) (1996), 2093-2167, MR 97f:20018, Zbl 901.20031.

[19] J.S. Williams, *Prime graph components of finite groups*, J. Algebra, **69** (1981), 485-513, MR 82j:20054, Zbl 471.20013.

Received April 28, 1999 and revised August 23, 1999. The first author was partially supported by BSF 92-003200 and by grant no. 6782-1-95 from the Israeli Ministry of Science and Art. The second author was partially supported by an NSF grant.

Ben-Gurion University
Beer-Sheva 84105, Israel
E-mail address: yoavs@math.bgu.ac.il

University of Oregon
Eugene, OR 97403-122
E-mail address: seitz@math.uoregon.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 202, No. 1, 2002

HARMONIC MAPS FROM $\mathbb{R}^n$ TO $\mathbb{H}^m$ WITH SYMMETRY

Yuguang Shi and Luen-Fai Tam

It is known that there is no nonconstant bounded harmonic map from the Euclidean space $\mathbb{R}^n$ to the hyperbolic space $\mathbb{H}^m$. This is a particular case of a result of S.-Y. Cheng. However, there are many polynomial growth harmonic maps from $\mathbb{R}^2$ to $\mathbb{H}^2$ by the results of Z. Han, L.-F. Tam, A. Treibergs and T. Wan. One of the purposes of this paper is to construct harmonic maps from $\mathbb{R}^n$ to $\mathbb{H}^m$ by prescribing boundary data at infinity. The boundary data is assumed to satisfy some symmetric properties. On the other hand, it was proved by Han-Tam-Treibergs-Wan that under some reasonable assumptions, the image of a harmonic diffeomorphism from $\mathbb{R}^2$ into $\mathbb{H}^2$ is an ideal polygon with $n+2$ vertices on the geometric boundary of $\mathbb{H}^2$ if and only if its Hopf differential is of the form ϕdz^2 where ϕ is a polynomial of degree n. It is unclear whether one can find explicit relation between the coefficients of ϕ and the vertices of the image of the harmonic map. The second purpose of this paper is to investigate this problem. We will explicitly demonstrate some families of polynomial holomorphic quadratic differentials, such that the harmonic maps from $\mathbb{R}^2$ into $\mathbb{H}^2$ with Hopf differentials in the same family will have the same image. In proving this, we first study the asymptotic behaviors of harmonic maps from $\mathbb{R}^2$ into $\mathbb{H}^2$ with polynomial Hopf differentials ϕdz^2. The result may have independent interest.

0. Introduction.

Let $\mathbb{R}^n$ be the Euclidean space, and $\mathbb{H}^n$ be the hyperbolic space. In [**HTTW**], it was proved that under some reasonable assumptions, the image of a harmonic diffeomorphism from $\mathbb{R}^2$ into $\mathbb{H}^2$ is an ideal polygon with $n+2$ vertices on the geometric boundary of $\mathbb{H}^2$ if and only if its Hopf differential is of the form ϕdz^2 where ϕ is a polynomial of degree n. Note that ϕ is a polynomial of degree n if and only if the harmonic map is of polynomial growth of order $\frac{n}{2}+1$, see [**TW**] for example. In [**LW**], it is shown that the closure of the image of a harmonic map from $\mathbb{R}^n$ into $\mathbb{H}^m$ with polynomial growth of order l will intersect the geometric boundary of $\mathbb{H}^m$ at no more than Cl^{n-1} points, where C is a constant independent of l. Moreover, the image lies in the convex hull of these points. In higher dimensions, unlike harmonic maps from hyperbolic space to hyperbolic space, there are very few examples of nontrivial harmonic maps from $\mathbb{R}^n$ into $\mathbb{H}^m$. In fact, if the image of a harmonic map from $\mathbb{R}^n$ to $\mathbb{H}^m$ is bounded, then the harmonic map must be constant [**Cg**]. Also, there is no rotationally symmetric harmonic map from $\mathbb{R}^n$ into $\mathbb{H}^n$ [**T**]. On the other hand, in [**WA**], (see also [**TW**]), it was shown that orientation preserving harmonic diffeomorphisms from $\mathbb{R}^2$ into $\mathbb{H}^2$ can be parametrized by their Hopf differentials, provided that the harmonic diffeomorphisms satisfy some natural conditions. In particular, one can construct harmonic diffeomorphisms from $\mathbb{R}^2$ to $\mathbb{H}^2$ with prescribed Hopf differentials. In [**HTTW**], harmonic diffeomorphisms with prescribed images had been constructed via the Gauss maps of constant mean curvature cuts in Minkowski three space. Both methods of constructions cannot be applied to higher dimensions. In this paper, we will use a more direct method to construct harmonic maps from $\mathbb{R}^n$ to $\mathbb{H}^m$ with prescribed boundary data at infinity. The boundary data is assumed to satisfy some symmetric properties. It should be remarked that if u is a harmonic map from $\mathbb{R}^2$ into $\mathbb{H}^{m-1}$ then u can be considered as a harmonic map from $\mathbb{R}^2$ into $\mathbb{H}^m$ by embedding $\mathbb{H}^{m-1}$ into $\mathbb{H}^m$. Also if u is a harmonic map from $\mathbb{R}^{n-1}$ into $\mathbb{H}^m$, then the map v from $\mathbb{R}^n = \mathbb{R}^{n-1} \times \mathbb{R}$ into $\mathbb{H}^m$ defined by $v(x,t) = u(x)$ for $(x,t) \in \mathbb{R}^{n-1} \times \mathbb{R}$ is harmonic. The harmonic maps we are going to construct are not in these categories, and are said to be nontrivial. Each of the constructed harmonic maps has polynomial growth, and the closure of its image in $\mathbb{H}^m \cup \partial\mathbb{H}^m$ intersects $\partial\mathbb{H}^m$ at finitely many points, where $\partial\mathbb{H}^m$ is the geometric boundary of $\mathbb{H}^m$. This can be considered as the first step to understand boundary value problem for harmonic maps from $\mathbb{R}^n$ into $\mathbb{H}^m$. The idea of construction is to find an approximate initial map with symmetry. Using the symmetry of the initial map, one can construct a harmonic map by compact exhaustion. The resulting harmonic map will be of bounded distance from the initial map.

In [**HTTW**], it was proved that if u is a harmonic diffeomorphism from $\mathbb{R}^2$ onto an ideal polygon with m vertices on $\partial\mathbb{H}^2$, then its Hopf differential is ϕdz^2 with ϕ to be a polynomial of degree $m-2$. However, it is unclear whether it is possible to find explicit relation between the coefficients of ϕ and these m points. The second purpose of this paper is to investigate this problem. We will explicitly demonstrate some families of polynomial holomorphic quadratic differentials, such that the harmonic maps from $\mathbb{R}^2$ into $\mathbb{H}^2$ with Hopf differentials in the same family will have the same image. In proving this, one needs to study asymptotic behaviors of harmonic maps from $\mathbb{R}^2$ into $\mathbb{H}^2$ with polynomial Hopf differentials. Some results in this direction had been obtained in [**HTTW**], using the techniques introduced in [**Wf**] and [**My**]. We will prove that if ϕ is of degree n, then there are $n+2$ rays, with equal angle between them, so that if u is a an orientation preserving harmonic map from $\mathbb{R}^2$ into $\mathbb{H}^2$ with Hopf differential ϕdz^2, then $u(z)$ will tend to infinity as $z \to \infty$ at the same rate along these rays. The result has its own interest and may be useful in the construction of harmonic maps $\mathbb{R}^2$ into $\mathbb{H}^2$ with prescribed data at infinity.

The structure of the paper is as follows. In §1, we will construct harmonic maps with symmetry from $\mathbb{R}^2$ to $\mathbb{H}^2$. In §2, we will use induction to construct nontrivial harmonic maps from $\mathbb{R}^2$ to $\mathbb{H}^m$, $m \geq 3$, and in §3, we will construct nontrivial harmonic maps from $\mathbb{R}^m$ to $\mathbb{H}^m$. In §4, we will study asymptotic behaviors of harmonic maps. In §5, we obtain some partial results on the explicit relation between the Hopf differential and the image of a harmonic map.

1. Harmonic maps from $\mathbb{R}^2$ to $\mathbb{H}^2$.

It was proved in [**WA**] (see also [**TW**]) that given a holomorphic quadratic differential $\phi(z)dz^2$ on $\mathbb{C}$, one can find a harmonic diffeomorphism from $\mathbb{C}$ into $\mathbb{H}^2$ such that the Hopf differential of the harmonic map is $\phi(z)dz^2$. Under certain conditions, the harmonic map is essentially unique. In particular, if $\phi(z) = z^m$, $m \geq 1$, using the result in [**HTTW**], one should be able to prove that up to an isometry of $\mathbb{H}^2$, the image is a regular ideal polygon of $m+2$ sides, see §5 for details. However, the method cannot be applied to higher dimensions. In this section, we will use another method to construct such harmonic maps. Using similar methods we will construct nontrivial harmonic maps with symmetry from $\mathbb{R}^2$ into $\mathbb{H}^m$, and $\mathbb{R}^m$ into $\mathbb{H}^m$, with $m \geq 2$ in the next two sections.

Let $n \geq 3$ be an integer. In $\mathbb{R}^2$, using polar coordinates the harmonic function

$$f(z) = f(re^{\sqrt{-1}\theta}) = r^{\frac{n}{2}} \sin\left(\frac{n}{2}\theta\right)$$

is zero on the rays $\theta = \theta_k$, where $0 \leq k \leq n-1$, where $\theta_k = \frac{2k\pi}{n}$, and $|f|$ is positive on $\theta_k < \theta < \theta_{k+1}$. Note that the ray $\theta = \theta_0$ is the same as the ray $\theta = \theta_n$. For each k, let W_k be the wedge defined by $\theta_k \leq \theta \leq \theta_{k+1}$.

Let us use the Poincaré disk model for $\mathbb{H}^2$. Let $a_k = e^{\frac{(2k+1)\pi\sqrt{-1}}{n}}$, $k = 0, ..., n-1$, which are identified as points on the geometric boundary of $\mathbb{H}^2$. Let o be the origin of the unit disk $\mathbb{D}$, and let γ_k be the geodesic from o to a_k in $\mathbb{H}^2$, parametrized by arc length. Define a map $g : \mathbb{R}^2 \to \mathbb{H}^2$ as follows. In the wedge $\theta_k \leq \theta \leq \theta_{k+1}$, let

$$g(z) = \gamma_k(|f(z)|).$$

Since $f = 0$ on each ray $\{\theta = \theta_k\}$, g is well-defined. g satisfies the following properties:

(i) g is a Lipschitz map, which is smooth and harmonic in the interior of each wedge W_k.

(ii) For any $z \in \mathbb{C}$, $g(e^{2\sqrt{-1}\theta_k}\bar{z}) = e^{2\sqrt{-1}\theta_k}\overline{g(z)}$.
(iii) $g(e^{\sqrt{-1}\theta_1}z) = e^{\sqrt{-1}\theta_1}g(z)$.

Lemma 1.1. *For any $R > 0$, let u_R be the harmonic map from $B(R)$ into $\mathbb{H}^2$, where $B(R)$ is the disk of radius R with center at the origin in $\mathbb{R}^2$, such that $u_R = g$ on $\partial B(R)$. Then there is a constant C_1 which is independent of R, such that*

$$d(u_R(z), g(z)) \le C_1$$

for all R and for all $z \in B(R)$.

Proof. By (iii) and the uniqueness of harmonic maps, we have

$$u_R(e^{\sqrt{-1}\theta_1}z) = e^{\sqrt{-1}\theta_1}u_R(z).$$

Hence it is sufficient to prove that

$$d(u_R(z), g(z)) \le C_1$$

for all $z \in B(R) \cap W_0$, where W_0 is the wedge defined above. By the definition of u_R,

$$u_R(z) = g(z) \tag{1.1}$$

for $z \in \partial B(R) \cap W_0$. We want to show that $d(u_R(z), g(z))$ is bounded on $\partial W_0 \cap B(R)$ by a constant independent of R. Since W_0 is bounded by two rays $\theta = \theta_0$, $\theta = \theta_1$, by symmetry it is sufficient to prove that $d(u_R(z), g(z))$ is uniformly bounded on $\{\theta = \theta_0\} \cap B(R)$. By (ii), $g(z) = \overline{g(\bar{z})}$. Hence by the uniqueness theorem on harmonic maps, we have $u_R(z) = \overline{u_R(\bar{z})}$. This implies that $u_R(z)$ lies on the real axis, for all $z \in \{\theta = \theta_0\} \cap B(R)$. Observe that the image of u_R lies inside the convex hull A of the ideal boundary points a_k, $0 \le k \le n-1$, and the closure of A in $\mathbb{H}^2 \cup \partial\mathbb{H}^2$ intersects $\partial\mathbb{H}^2$ at the points a_k. Suppose n is even, then no a_k is on the real axis. Hence there is a constant C_2 independent of R, such that

$$\begin{aligned} d(u_R(z), g(z)) &= d(u_R(z), o) \\ &\le C_2 \end{aligned} \tag{1.2}$$

for all $z \in \{\theta = \theta_0\} \cap B(R)$, see Figure 1. Suppose n is odd, we want to show that $u_R(z)$ lies on the positive real axis, for all $z \in \{\theta = \theta_0\} \cap B(R)$. This will imply that (1.2) is still true in this case, because no a_k is on the positive real axis. By the definition of g, we see that g maps the upper half space into the that part of $\mathbb{D}^2$ which lies on the upper half space. Since $u_R(z)$ lies on the real axis if z is real, u_R also maps the upper half space into the that part of $\mathbb{D}^2$ which lies on the upper half space. One can prove similarly that u_R maps the half space bounded by the rays $\theta = \frac{2\pi}{n}$ and $\theta = \pi + \frac{2\pi}{n}$ which containing the positive real axis into the same half space, see Figure 2. In particular, $u_R(z)$ lies on the positive real axis, for all $z \in \{\theta = \theta_0\} \cap B(R)$. So

(1.2) is true for all $z \in \partial(B(R) \cap W_0)$. Since $d(u_R(z), g(z))$ is subharmonic, the lemma follows from the maximum principle.

By Lemma 1.1, passing to a subsequence if necessary, u_R will converge to a harmonic map u such that $d(u(z), g(z))$ is uniformly bounded. In fact, u is a diffeomorphism. We can prove this fact as follows. For each $R > 0$, let us construct a harmonic map v_R from $B(R)$ into $\mathbb{H}^2$ in the following way. Let $b_k = \gamma_k(R^{\frac{n}{2}})$ and let β_k be the minimal geodesic joining b_k to b_{k-1}. Let α_k be the minimal geodesic joining a_k to a_{k-1}. It is easy to see that the distance from a point on γ_k or γ_{k-1} to α_k is bounded by a constant C_3 which is independent of R. Define a map Π_k from $\gamma_k|[0, b_k]$ and $\gamma_{k-1}|[0, b_{k-1}]$ into the line containing β_k, by nearest point projection. Then

$$d(\gamma_k(s), \Pi_k(\gamma_k(s))) \leq C_1. \tag{1.3}$$

Π_k is surjective and is continuous. Let v_R be the harmonic map from $B(R)$ into $\mathbb{H}^2$, such that on the $\partial B(R) \cap W_k$ $v_R(z) = \Pi_k(g(z))$. Note that the boundary map is a homeomorphism from $\partial B(R)$ onto the boundary of the geodesic polygon with boundary $\cup_k \beta_k$. Here are some properties of v_R. By [**SY**], we have:

Lemma 1.2. *v_R is a diffeomorphism onto its image.*

By Lemma 1.1, and (1.3), there is a constant C_4 which is independent of R such that

$$\sup_{x \in B(R)} d(v_R(z), g(z)) \leq C_4.$$

Hence, passing to a subsequence, v_R converge to a harmonic map v, such that

$$d(v(z), g(z)) \leq C_4. \tag{1.4}$$

Lemma 1.3. *Let ϕdz^2 be the Hopf differential of v. Then ϕ is a polynomial of degree $n - 2$.*

Proof. By the construction,

$$d(o, g(z)) \leq |z|^{\frac{n}{2}}.$$

By (1.4), we see that

$$d(o, v(z)) \leq C_4 + |z|^{\frac{n}{2}}.$$

By the energy density estimate [**Cg**], there is a constant C_5 independent of z such that

$$e(u)(z) \leq C_5(|z|^{n-2} + 1).$$

Since $|\phi|(z) \leq e(v)$, we conclude that ϕ is a polynomial of degree at most $n - 2$. Suppose the degree of ϕ is less than or equal to $n - 3$. Let $\phi_R dz^2$ be the Hopf differential of v_R. Then given any $R_0 > 0$ there is R_1 such that if $R > R_1$, then

$$|\phi_R(z)| \leq C_6(|z|^{n-3} + 1),$$

in $B(R_0)$ for some constant C_6 which is independent of R_0, where ϕ_R is the Hopf differential of v_R. Using an argument of [**TW**], we conclude that in $B(\frac{R_0}{2})$,

$$e(v_R)(z) \le C_7(|z|^{n-3}+1)$$

for some constant C_7 independent of R_0, if R is large enough. Let $R \to \infty$, and then let $R_0 \to \infty$, we have

$$e(v)(z) \le C_7(|z|^{n-3}+1).$$

This would imply

$$d(o, v(z)) \le C_8(|z|^{(n-1)/2}+1)$$

for some constant C_8. By (1.4), and the definition of g, this is impossible. Hence the degree of ϕ must be $n-2$.

Lemma 1.4. *v is a diffeomorphism onto its image.*

Proof. Since the Jacobian J_R of v_R is positive in $B(R)$, the Jacobian J of v satisfies $J \ge 0$. First we want to show that $J > 0$ somewhere. Suppose not, then $J \equiv 0$. Since $J = ||\partial v||^2 - ||\bar\partial v||^2$, where $||\partial v|| = \sigma|\frac{\partial v}{\partial z}|$, and $||\bar\partial v|| = \sigma|\frac{\partial v}{\partial \bar z}|$, $\sigma^2|dv|^2$ is the metric on $\mathbb{H}^2$. we have

$$||\partial v||^2 \equiv ||\bar\partial v||^2.$$

On the other hand,

$$|\phi|^2 = ||\partial v||^2 \cdot ||\bar\partial v||^2.$$

We have

$$|\phi| = ||\partial v||^2.$$

Since ϕ is a polynomial of degree $n-2$, there is $R_0 > 0$ such that all the zeros of ϕ lies inside $B(\frac{R_0}{2})$. For each R, $||\partial v_R|| > 0$, and let $w_R = \log ||\partial v_R||$. Then

$$\Delta w_R = J_R(u_R).$$

We have

$$\int_{\partial B(R_0)} \frac{\partial w_R}{\partial r} = \int_{B(R_0)} \Delta w_R$$
$$= \int_{B(R_0)} J_R.$$

Since on $\partial B(R_0)$, $||\partial v||^2 = |\phi| > 0$, let $R \to \infty$, we have

$$\int_{\partial B(R_0)} \frac{\partial w}{\partial r} = \int_{B(R_0)} J.$$

However, $w = \frac{1}{2}\log|\phi|$, and the degree of ϕ is at least 1, moreover, all zeros of ϕ lie inside $B(R_0)$, we conclude that

$$\int_{B(R_0)} J > 0. \tag{1.5}$$

Hence $J > 0$ somewhere, and (1.5) is true for some $R_0 > 0$. This implies that there is $\delta > 0$ such that if R is large then

$$\int_{B(R_0)} J_R \geq \delta. \tag{1.6}$$

Apply Theorem 7.1 in [**J**] to each map v_R, we conclude that for any R_1, there is $\epsilon > 0$, such that

$$J_R(v) \geq \epsilon > 0$$

in $B(R_1)$ provided R is large enough. This implies $J(v) > 0$ everywhere and v is a diffeomorphism onto its image.

Since $d(v(z), u(z))$ is uniformly bounded and subharmonic, $d(v(z), u(z))$ is a constant function. It is easy to see that $v(0) = u(0)$, and so $u \equiv v$. On the other hand, since $||\partial u||^2 \geq |\phi|$ and ϕ is a polynomial, we see that $||\partial u||^2 dz^2$ is complete. By the result of [**HTTW**], the image of u is a ideal polygon of n sides and so the image of u is the polygon spanned by the $a_k' s$, and we have the following:

Theorem 1.5. *Let $n \geq 3$, and let $a_k = e^{\frac{(2k+1)\pi\sqrt{-1}}{n}}$, $k = 0, \ldots, n-1$. Then there is a harmonic diffeomorphism u from $\mathbb{R}^2$ into $\mathbb{H}^2$ whose image is the ideal polygon spanned by the a_k's. Moreover, u satisfies*

$$u(e^{2\sqrt{-1}\theta_k}\bar{z}) = e^{2\sqrt{-1}\theta_k}\overline{u(z)}$$

and

$$u(e^{i\theta_1}z) = e^{\sqrt{-1}\theta_1}u(z).$$

In case of $n = 4$, we can do more. Let a_k, $1 \leq k \leq 4$ be four points on the unit circle, such that they are the vertices of a rectangle which is symmetric with respect to the real and imaginary axes.

Proposition 1.6. *There is a harmonic diffeomorphism from $\mathbb{R}^2$ into $\mathbb{H}^2$ whose image is the ideal polygon spanned by the a_k's. Moreover, u satisfies*

$$u(\bar{z}) = \overline{u(z)},$$

and

$$u(-\bar{z}) = -\overline{u(z)}.$$

The proof is similar to the proof of Theorem 1.5. We should remark that for any four points on the unit circle, there is a conformal map of the unit disk, which carries these four points to some a_k's satisfying the condition of Proposition 1.6.

2. Harmonic maps from $\mathbb{R}^2$ into $\mathbb{H}^m$.

In this section, we will use the harmonic maps constructed in §1 to obtain harmonic maps from $\mathbb{R}^2 = \mathbb{C}$ into $\mathbb{H}^m$, which are nontrivial in the sense that the image of each of the maps is not contained in any nontrivial totally geodesic submanifold in $\mathbb{H}^m$. We always use the Poincaré unit ball model for $\mathbb{H}^m$. Namely, $\mathbb{H}^m$ is identified with the unit ball $\mathbb{B}^m$ in $\mathbb{R}^m$ with the Poincaré metric, and the geometric boundary $\partial\mathbb{H}^m$ is identified with the unit sphere $\mathbb{S}^{m-1}$. For any set A in $\mathbb{H}^m \cup \partial\mathbb{H}^m$, we denote $\overline{A}$ to be the closure of A in $\mathbb{H}^m \cup \partial\mathbb{H}^m$, and denote the convex hull of A by $\mathrm{Con}\,(A)$. We will use the following fact: Suppose A is a close set in $\mathbb{H}^m \cup \partial\mathbb{H}^m$, then $\overline{\mathrm{Con}\,(A)} \cap \partial\mathbb{H}^m = A \cap \partial\mathbb{H}^m$.

Let $n \geq 4$ be an even number. Let $\theta_k = \frac{2k\pi}{n}$ and let W_k be the wedge in $\mathbb{R}^2$ defined by $\theta_k \leq \theta \leq \theta_{k+1}$ in polar coordinates. Note that $\theta_k = k\theta_1$. By Theorem 1.5, we can find a harmonic diffeomorphism u from $\mathbb{C} = \mathbb{R}^2$ into $\mathbb{H}^2$, such that:

(a) In the Poincaré disk model of $\mathbb{H}^2$, if we write

$$u(z) = (u^1(z), u^2(z)),$$

then $u^1(z) = 0$ on $\Im(z) = 0$, where $\Im(z)$ is the imaginary part of z;

(b) $\overline{u(\mathbb{R}^2)} \cap \partial\mathbb{H}^2$ does not contain the points $(0, \pm 1)$.

From (a) and (b), we have

(c) $\sup_{z\in\mathbb{R}^2,\ \Im(z)=0} d(u(z), 0) < \infty$.

From (b), we also have:

(b′) If $(a^1, a^2) \in \overline{u(\mathbb{R}^2)} \cap \partial\mathbb{H}^2$, then $a^1 \neq 0$.

We are going to use u to construct a harmonic map from $\mathbb{R}^2$ into $\mathbb{H}^3$. Identify $\mathbb{H}^2$ with $\{(v^1, v^2, v^3) \in \mathbb{H}^3 |\ v^2 = 0\}$. Then $u : \mathbb{R}^2 \to \mathbb{H}^2 \subset \mathbb{H}^3$ is also harmonic, and

$$u(z) = (u^1(z), 0, u^2(z)). \tag{2.1}$$

Define a harmonic map v from W_0 into $\mathbb{H}^3$ in the following way, see Figure 3. Let

$$\Psi : \{z \in \mathbb{C}|\ \Im(z) > 0\} \to \text{ interior of } W_0,$$

be a conformal diffeomorphism, $\Psi(\{\Im(z) = 0\}) = \partial W_0$ and Ψ is homeomorphism between $\Im(z) \geq 0$ and W_0 . Let $v(z) = u \circ \Psi^{-1}(z)$. Then v is a harmonic map from W_0 into $\mathbb{H}^3$, such that:

(i) $v(z) = (v^1(z), 0, v^3(z))$;

(ii) $v(z) = (0, 0, v^3(z))$ for $z \in \partial W_0$;

(iii) $\sup_{z\in\partial W_0} d(v(z), 0) < \infty$;

(iv) v is continuous up to the boundary of W_0

(v) suppose $(a^1, a^2, a^3) \in \overline{v(W_0)} \cap \partial\mathbb{H}^3$, $a^1 \neq 0$.

Property (v) follows from property (b′) of u and the fact that if $(a^1, a^2, a^3) \in \overline{v(W_0)} \cap \partial\mathbb{H}^3$ then $a^2 = 0$ and $(a^1, a^3) \in \overline{u(\mathbb{R}^2)} \cap \partial\mathbb{H}^2$.

Define g as follows, see Figure 4. Let us write any point $v = (v^1, v^2, v^3)$ of $\mathbb{H}^3$ in the form $(v^1 + \sqrt{-1}v^2, v^3)$. Let $g(z) = v(z)$ for $z \in W_0$. Suppose we have defined $g = (g^1, g^2, g^3) = (g^1 + \sqrt{-1}g^2, g^3)$ on W_k, $0 \le k < n-1$, then for $z \in W_{k+1}$, let

$$g(z) = \left(e^{2\sqrt{-1}(\theta_{k+1} - \frac{\pi}{n})}(g^1 - \sqrt{-1}g^2)(\hat{z}), g^3(\hat{z})\right) \\ = \left(e^{2\sqrt{-1}(k+\frac{1}{2})\theta_1}(g^1 - \sqrt{-1}g^2)(\hat{z}), g^3(\hat{z})\right) \tag{2.2}$$

here $\hat{z} = e^{2\sqrt{-1}\theta_{k+1}}\bar{z}$ which is in W_k. Here we simply 'reflect' g along the ray $\theta = \theta_{k+1}$ in the domain, and $\theta = \theta_{k+1} - \frac{\pi}{n} = (k + \frac{1}{2})\theta_1$ in the target. Then g is harmonic on the interior of each W_k. Suppose n is even, then g^3 is a well-defined and continuous function on $\mathbb{R}^2$, and since $g = (0, 0, g^3)$ on ∂W_k for all k, g is well-defined and continuous.

Lemma 2.1. *Suppose n is even, and n is not a multiple of* 4. *Then the map g defined above satisfies:*

(i) $g(z) = (e^{2\sqrt{-1}(k-\frac{1}{2})\theta_1}(g^1 - \sqrt{-1}g^2)(\hat{z}), g^3(\hat{z}))$, *where* $\hat{z} = e^{2\sqrt{-1}\theta_k}\bar{z}$, *for all z and* $0 \le k \le n-1$;

(ii) $\sup_{z \in \partial W_k} d(g(z), 0) < \infty$, *for* $0 \le k \le n-1$; *and*

(iii) *suppose* $(a^1, a^2, a^3) \in \overline{g(\mathbb{R}^2)} \cap \partial\mathbb{H}^3$, *then* $a^1 \neq 0$, *and* $\arg(a^1 + \sqrt{-1}a^2) = \theta_k$ *or* $\theta_k + \pi$, *for some* $0 \le k \le n-1$.

Proof. Let $z_0 \in W_0$, define z_s inductively by

$$z_{s+1} = e^{2\sqrt{-1}(s+1)\theta_1}\bar{z}_s,$$

for $s = 0, \ldots, n-1$. Then $z_s \in W_s$. Suppose $s = 2l$, then

$$z_s = e^{2\sqrt{-1}l\theta_1}z_0 \tag{2.3}$$

and

$$g(z_s) = (e^{2\sqrt{-1}l\theta_1}(g^1 + \sqrt{-1}g^2)(z_0), g^3(z_0)) \\ = (e^{\sqrt{-1}s\theta_1}(g^1 + \sqrt{-1}g^2)(z_0), g^3(z_0)). \tag{2.4}$$

If $s = 2l + 1$, then

$$z_s = \hat{z}_{2l} \\ = e^{2\sqrt{-1}s\theta_1}\bar{z}_{2l} \\ = e^{2\sqrt{-1}(l+1)\theta_1}\bar{z}_0, \tag{2.5}$$

and

$$g(z_s) = g(\hat{z}_{2l}) = \left(e^{2\sqrt{-1}(s-\frac{1}{2})\theta_1}(g^1 - \sqrt{-1}g^2)(z_{2l}), g^3(z_{2l})\right) = \left(e^{2\sqrt{-1}(l+\frac{1}{2})\theta_1}(g^1 - \sqrt{-1}g^2)(z_0), g^3(z_0)\right) = \left(e^{\sqrt{-1}s\theta_1}(g^1 - \sqrt{-1}g^2)(z_0), g^3(z_0)\right). \tag{2.6}$$

Hence $z_n = z_0$, and $z_{n-1} = \bar{z}_0$, because n is even, and

$$g(\bar{z}_0) = (e^{-\sqrt{-1}\theta_1}(g^1 - \sqrt{-1}g^2)(z_0), g^3(z_0)). \tag{2.7}$$

Now suppose $z = \rho e^{i\alpha}$ where $\theta_m \le \alpha < \theta_{m+1}$ for some $0 \le m \le n-1$. Then there exists $z_0 = \rho e^{i\alpha_0}$ with $0 \le \alpha_0 < \theta_1$, such that $z_m = z$. If $m = 2p$, then

$$\hat{z} = e^{2\sqrt{-1}\theta_k}\bar{z} = e^{2\sqrt{-1}(k-p)\theta_1}\bar{z}_0.$$

Without loss of generality, we may assume that $0 \le 2(k-p) \le n-1$. If $k - p = 0$, then, apply (2.4) to $g(\bar{z}_0) = g(\hat{z})$ and (2.7) to $g(\hat{z})$, we have

$$g(z) = \left(e^{2\sqrt{-1}p\theta_1}(g^1 + \sqrt{-1}g^2)(z_0), g^3(z_0)\right) = \left(e^{2\sqrt{-1}k\theta_1}(g^1 + \sqrt{-1}g^2)(z_0), g^3(z_0)\right) = \left(e^{2\sqrt{-1}(k-\frac{1}{2})\theta_1}(g^1 - \sqrt{-1}g^2)(\bar{z}_0), g^3(\bar{z}_0)\right) = \left(e^{2\sqrt{-1}(k-\frac{1}{2})\theta_1}(g^1 - \sqrt{-1}g^2)(\hat{z}), g^3(\hat{z})\right).$$

So (i) is true in this case. Suppose $k - p = l + 1$, with $l \ge 0$, then we can apply (2.5) and (2.6)

$$g(z) = \left(e^{2\sqrt{-1}p\theta_1}(g^1(z_0) + \sqrt{-1}g^2(z_0)), g^3(z_0)\right) = \left(e^{2\sqrt{-1}(p+k-p-1+\frac{1}{2})\theta_1}(g^1(\hat{z}) - \sqrt{-1}g^2(\hat{z})), g^3(\hat{z})\right) = \left(e^{2\sqrt{-1}(k-\frac{1}{2})\theta_1}(g^1(\hat{z}) - \sqrt{-1}g^2(\hat{z})), g^3(\hat{z})\right).$$

Then (i) is still true. The case that $m = 2p + 1$ can be proved similarly. The proof of (i) is completed. (ii) can be derived from the definition of g and property (iii) of v. To prove (iii), let $(a^1, a^2, a^3) \in \overline{g(\mathbb{R}^2)} \cap \partial\mathbb{H}^3$, then $(a^1, a^2, a^3) \in \overline{g(W_k)} \cap \partial\mathbb{H}^3$, for some $0 \le k \le n-1$. Since $g = v$ on W_0, by the definition of v and property (v) of v, if $(a^1, a^2, a^3) \in \overline{g(W_0)} \cap \partial\mathbb{H}^3$, then $a^2 = 0$, and $a^1 \ne 0$. In particular, $\arg(a^1 + \sqrt{-1}a^2) = \theta_0 = 0$ or π. Now suppose $(a^1, a^2, a^3) \in \overline{g(W_k)} \cap \partial\mathbb{H}^3$, for $1 \le k \le n-1$, then by (2.4), and (2.6), there is $(b^1, 0, b^3) \in \overline{g(W_0)} \cap \partial\mathbb{H}^3$, such that

$$a^1 + \sqrt{-1}a^2 = e^{\sqrt{-1}\theta_k} b^1.$$

Since n is not a multiple of 4, $e^{\sqrt{-1}\theta_k} \neq \pm i$, and since $b^1 \neq 0$ and is real, we have $a^1 \neq 0$. Moreover, $\arg(a^1 + \sqrt{-1}a^2) = \theta_k$ or $\theta_k + \pi$.

Theorem 2.2. *Let n and $g(z)$ be as Lemma 2.1. There exists a harmonic map h from $\mathbb{R}^2$ into $\mathbb{H}^3$, such that*

$$\sup_{z\in\mathbb{C}} d(h(z), g(z)) < \infty.$$

Moreover:

(a) *In the Poincaré ball model of $\mathbb{H}^3$, if $\Im(z) = 0$ and if we let*

$$h(z) = (h^1(z), h^2(z), h^3(z)),$$

then $\arg(h^1 + \sqrt{-1}h^2)(z) = -\frac{1}{2}\theta_1$ *or* $\pi - \frac{1}{2}\theta_1$*;*

(b) *suppose $(a^1, a^2, a^3) \in \overline{h(\mathbb{R}^2)} \cap \partial\mathbb{H}^3$, then $a^1 \neq 0$; and*

(c) $\sup_{z\in\mathbb{R}^2,\ \Im(z)=0} d(h(z), 0) < \infty$.

If, in addition, $\overline{u(\{\Im(z) \ge 0\})} \cap \partial\mathbb{H}^2$ is not contained in any straight line in the plane, then $\overline{h(\{\Im(z) \ge 0\})} \cap \partial\mathbb{H}^3$ is not contained in any hyperplane in $\mathbb{R}^3$. In particular, the image of h is not contained in any totally geodesic submanifold of dimension 2 in $\mathbb{H}^3$.

Proof. For any $R > 0$, let B_R be the disk of radius R with center at the origin in $\mathbb{R}^2$. Let h_R be the harmonic map from B_R into $\mathbb{H}^3$, such that $h_R = g$ on ∂B_R. If we write $h_R = (h_R^1, h_R^2, h_R^3) = (h_R^1 + \sqrt{-1}h_R^2, h_R^3)$, then by the uniqueness of harmonic maps and Lemma 2.1, we have

$$h_R(z) = \left(e^{2\sqrt{-1}(k-\frac{1}{2})\theta_1}(h_R^1 - \sqrt{-1}h_R^2)(\hat{z}), h_R^3(\hat{z})\right) \tag{2.8}$$

for any $z \in B_R$, where $\hat{z} = e^{2\sqrt{-1}\theta_k}\bar{z}$, $0 \le k \le n-1$. We want to show that there exists a constant C_1 independent of R such that

$$d(h_R(z), g(z)) \le C_1 \tag{2.9}$$

for all $z \in B_R$. Obviously, we only have to prove that (2.9) is true for all $z \in W_k \cap B_R$, for all $0 \le k \le n-1$. Let us consider W_0 for example. $\partial(W_0 \cap B_R)$ is the union of $W_0 \cap \partial B_R$, $\{\theta = 0\} \cap B_R$, and $\{\theta = \theta_1\} \cap B_R$. On $W_0 \cap \partial B_R$, $h_R = g$. On the other hand, for $z \in \{\theta = 0\} \cap B_R$, we have $z = \bar{z}$, and so by (2.8) with $k = 0$,

$$\begin{aligned} h_R(z) &= (e^{-\sqrt{-1}\theta_1}(h_R^1 - \sqrt{-1}h_R^2)(\bar{z}), h_R^3(\bar{z})) \\ &= (e^{-\sqrt{-1}\theta_1}(h_R^1 - \sqrt{-1}h_R^2)(z), h_R^3(z)). \end{aligned}$$

Hence $h_R(z) \in \Pi$ where Π is the plane $(v^1, v,^2, v^3) \in \mathbb{H}^3$, such that $\arg(v^1 + \sqrt{-1}v^2) = -\frac{1}{2}\theta_1$ or $\pi - \frac{1}{2}\theta_1$. Similarly, if z is in $\{\theta = \pi\} \cap B_R$, then $h_R(z)$ is also in Π. On the other hand, it is well-known that $u_R(B_R)$ is contained in the convex hull of $u_R(\partial B_R)$, which in turn is contained in the convex hull of $g(\mathbb{R}^2)$. Since $\overline{\mathrm{Con}\,(g(\mathbb{R}^2))} \cap \partial\mathbb{H}^3 = \overline{g(\mathbb{R}^2)} \cap \partial\mathbb{H}^3$, by Lemma 2.1 (iii) we conclude that if $(a^1, a^2, a^3) \in \overline{\mathrm{Con}\,(g(\mathbb{R}^2))} \cap \partial\mathbb{H}^3$, then $\arg(a^1 + \sqrt{-1}a^2) = \theta_k$ for some k, and $a^1 \neq 0$. However, by the definition of Π, if (a^1, a^2, a^3) is also in Π, then $\arg(a^1 + \sqrt{-1}a^2) = -\frac{1}{2}\theta_1$ or $\pi - \frac{1}{2}\theta_1$, which are not equal to θ_k modulo a multiple of 2π, because n is even. So

$$\Pi \cap \overline{\mathrm{Con}\,(g(\mathbb{R}^2))} \cap \partial\mathbb{H}^3 = \emptyset.$$

Since $h_R(z) \in \Pi$ for $z \in \{\theta = 0\} \cap B_R$, there exists a constant C_2 independent of R such that

$$d(h_R(z), 0) \leq C_2 \tag{2.10}$$

for $z \in \{\theta = 0\} \cap B_R$. By Lemma 2.1, there exists a constant C_3 independent of R such that for all $z \in \{\theta = 0\} \cap B_R$

$$d(g(z), 0) \leq C_3.$$

Combine this with (2.10), we have

$$d(h_R(z), g(z)) \leq C_2 + C_3$$

for all $z \in \{\theta = 0\} \cap B_R$. Similarly, one can prove that

$$d(h_R(z), g(z)) \leq C_4$$

for some constant C_4 independent of R, for all $z \in \{\theta = \theta_1\} \cap B_R$. Since g is harmonic on W_0, $d(h_R(z), g(z))$ is subharmonic on W_0. By the maximum principle, (2.9) is true on W_0. Similarly, (2.9) is true on W_k, for all k. By (2.9), passing to a subsequence if necessary, let $R \to \infty$, h_R converge to a harmonic map h from $\mathbb{R}^2$ to $\mathbb{H}^3$, such that

$$\sup_{z \in \mathbb{R}^2} d(h(z), g(z)) \leq C_1$$

for some constant C_1. In particular, $\overline{h(\mathbb{R}^2)} \cap \partial\mathbb{H}^3 = \overline{g(\mathbb{R}^2)} \cap \partial\mathbb{H}^3$. From this and Lemma 2.1, (b) follows. (c) follows from (2.9) and the property (ii) of g in Lemma 2.1. Since each h_R satisfies (a), so does h. It is well-known that a totally geodesic submanifold M is contained in a sphere or a hyperplane which intersects $\mathbb{S}^2$ orthogonally, see [**Sk**] for example. This implies that $\overline{M} \cap \partial\mathbb{H}^3$ is contained in a hyperplane. Hence, to prove the last statement, let us suppose $\overline{u(\{\Im(z) \geq 0\})} \cap \partial\mathbb{H}^2$ is not contained in any straight line in the plane, then it is sufficient to prove that the intersection of the closure of the image of h with $\partial\mathbb{H}^3$ is not contained in a hyperplane. By the construction of g, $g(W_0)$ consists of those points $(u^1(z), 0, u^2(z))$ with $\Im(z) > 0$. So

$$\overline{g(W_0)} \cap \partial\mathbb{H}^3 = \left\{(v^1, 0, v^3) |\ (v^1, v^3) \in \overline{u(\{\Im(z) \geq 0\})} \cap \partial\mathbb{H}^2\right\},$$

and the smallest affine subspace of $\mathbb{R}^3$ which contains $\overline{g(W_0)} \cap \partial\mathbb{H}^3$ is the subspace defined by $v^2 = 0$. By the definition of g,

$$\overline{g(W_1)} \cap \partial\mathbb{H}^3 = \left\{ (e^{\sqrt{-1}\theta_1}(v^1 - \sqrt{-1}v^2), v^3) | \ (v^1, v^2, v^3) \in \overline{g(W_0)} \cap \partial\mathbb{H}^3 \right\}.$$

Since $\theta_1 = \frac{2\pi}{n}$, $\overline{g(W_1)} \cap \partial\mathbb{H}^3$ is not contained in the subspace $v^2 = 0$. Since $W_0 \cup W_1$ is contained in $\Im(z) \geq 0$, we conclude that $\overline{g(\{\Im(z) \geq 0\})} \cap \partial\mathbb{H}^3$ is not contained in any hyperplane of $\mathbb{R}^3$. Using the fact that $d(h(z), g(z))$ is uniformly bounded from above, the same is true for h. From this, the last statement of the theorem follows.

By composing h with the isometry

$$(v^1 + \sqrt{-1}v^2, v^3) \to \left(e^{\frac{\sqrt{-1}}{2}(\theta_1 + \pi)}(v^1 + \sqrt{-1}v^2), v^3 \right)$$

on $\mathbb{H}^3$, we obtain a harmonic map u. Obviously, u also satisfies (c) of Theorem 2.2, (with h replaced by u). Also $u^1(z) = 0$ on $\Im(z) = 0$. Suppose $(a^1, a^2, a^3) \in \overline{u(\mathbb{R}^2)} \cap \partial\mathbb{H}^3$, then $a^1 + \sqrt{-1}a^2 = e^{\frac{\sqrt{-1}}{2}(\theta_1 + \pi)}(b^1 + \sqrt{-1}b^2)$ for some $(b^1, b^2, b^3) \in \overline{h(\mathbb{R}^2)} \cap \partial\mathbb{H}^3$. From the proof we see that $b^1 + \sqrt{-1}b^2 = e^{\sqrt{-1}\theta_k}c$ for some $c \neq 0$, and for some $0 \leq k \leq n-1$. From this we conclude that $a^1 \neq 0$. Here we use the fact that n is even again.

We can proceed as before to use u to construct a harmonic map from $\mathbb{R}^2$ into $\mathbb{H}^4$. More precisely and more generally, suppose u is a harmonic map from $\mathbb{R}^2 \to \mathbb{H}^m$ for some $m \geq 2$, such that:

(a) In the Poincaré ball model of $\mathbb{H}^m$, if we write

$$u(z) = (u^1(z), u^2(z), \ldots, u^m(z)),$$

then $u^1(z) = 0$ on $\Im(z) = 0$;

(b) if $(a^1, \ldots, a^m) \in \overline{u(\mathbb{R}^2)} \cap \partial\mathbb{H}^m$ then $a^1 \neq 0$;

(c) $\sup_{z \in \mathbb{R}^2, \ \Im(z)=0} d(u(z), 0) < \infty$.

Let n be even, not divisible by 4, and defined θ_k, W_k, Ψ as before. Let $v(z) = u \circ \Psi^{-1}(z)$, for any $z \in W_0$. Define $g(z) = v(z)$ for any $z \in W_0$. Suppose we have already defined $g(z)$ on W_k, $0 \leq k \leq n-1$, then for any $z \in W_{k+1}$ define:

$$g(z) = \left(e^{2\sqrt{-1}(k+\frac{1}{2})\theta_1}(g^1 - \sqrt{-1}g^2)(\hat{z}), g^3(\hat{z}), \ldots, g^{m+1}(\hat{z}) \right)$$

here $\hat{z} = e^{2\sqrt{-1}\theta_{k+1}}\overline{z} \in W_k$. Using similar methods as in Theorem 2.2, we can prove:

Theorem 2.2′. *Let $g(z)$ be as above. There exists a harmonic map h from $\mathbb{R}^2$ into $\mathbb{H}^{m+1}$, such that*

$$\sup_{z \in \mathbb{C}} d(h(z), g(z)) < \infty.$$

Moreover:

(a) *In the Poincaré ball model of* $\mathbb{H}^{m+1}$, *if* $\Im(z) = 0$, *and if*

$$h(z) = (h^1(z), h^2(z), \ldots, h^{m+1}(z))$$

then $\arg(h^1 + \sqrt{-1}h^2)(z) = -\frac{1}{2}\theta_1$ *or* $\pi - \frac{1}{2}\theta_1$;

(b) *if* $(a^1, \ldots, a^{m+1}) \in \overline{h(\mathbb{R}^2)} \cap \partial\mathbb{H}^{m+1}$, *then* $a^1 \neq 0$; *and*

(c) $\sup_{z\in\mathbb{R}^2,\ \Im(z)=0} d(h(z), 0) < \infty$.

If, in addition, $\overline{u(\{\Im(z) \geq 0\})} \cap \partial\mathbb{H}^m$ *is not contained in any hyperplane in* $\mathbb{R}^m$, *then* $\overline{h(\{\Im(z) \geq 0\})} \cap \partial\mathbb{H}^{m+1}$ *is not contained in any hyperplane in* $\mathbb{R}^{m+1}$. *In particular, the image of* h *is not contained in any totally geodesic submanifold of dimension* m *in* $\mathbb{H}^{m+1}$.

Again by composing h with the isometry

$$(v^1 + \sqrt{-1}v^2, v^3, \ldots, v^{m+1}) \to (e^{\frac{\sqrt{-1}}{2}(\theta_1+\pi)}(v^1 + \sqrt{-1}v^2), v^3, \ldots, v^{m+1}),$$

we obtain a harmonic map from $\mathbb{R}^2$ into $\mathbb{H}^{m+1}$ satisfying required properties for the induction on construction.

Remark 2.1. (i) By the result in §1, it is easy to see that there are many harmonic maps u from $\mathbb{R}^2$ into $\mathbb{H}^2$, which satisfy the conditions in Theorem 2.2.

(ii) If we begin with a harmonic map u constructed in §1, and obtain harmonic maps inductively using Theorem 2.2, and 2.2′, then the harmonic maps will be of polynomial growth, and the closure of the image of each of the maps intersects the geometric boundary of the hyperbolic space at finitely many points. This is related to the results in [**LW**].

3. Harmonic maps from $\mathbb{R}^m$ into $\mathbb{H}^m$.

In this section, we will use methods similar to those in §1 and §2 to construct nontrivial harmonic maps from $\mathbb{R}^m$ into $\mathbb{H}^m$, $m \geq 3$. First let us write $\mathbb{R}^m = \mathbb{R}^2 \times \mathbb{R}^{m-2}$. As in the previous section, let $n \geq 4$ be an even integer, $\theta_k = \frac{2k\pi}{n}$, $0 \leq k \leq n-1$, and let W_k be the wedge in $\mathbb{R}^2$ defined by $\theta_k \leq \theta \leq \theta_{k+1}$ in polar coordinates. Let $\Omega_k = W_k \times [0,\infty)^{m-2}$ which consists of points $(x^1, x^2, \ldots, x^m)$ with $(x^1, x^2) \in W_k$ and $x^j \geq 0$, for $3 \leq j \leq m$. We use the Poincaré unit ball model for $\mathbb{H}^m$ as before. Define a harmonic function f by

$$f(x^1, x^2, x^3, \ldots, x^m) = r^{\frac{n}{2}} \sin\left(\frac{n}{2}\theta\right) x^3 \cdots x^m,$$

on Ω_k, $k = 0, 1, \ldots, n-1$, where $x^1 + \sqrt{-1}x^2 = re^{\sqrt{-1}\theta}$. Let $\gamma : [0,\infty) \mapsto \mathbb{H}^m$ be the geodesic parametrized by arc length, such that $\gamma(0) = 0$,

$$\gamma(t) = (\gamma^1(t), 0, \gamma^3(t), \ldots, \gamma^m(t))$$

$\gamma^i(t) \geq 0$, and $\lim_{t\to\infty} \gamma^m(t) = ((m-1)^{-\frac{1}{2}}, 0, (m-1)^{-\frac{1}{2}}, \dots, (m-1)^{-\frac{1}{2}})$. Define $v : \Omega_0 \mapsto \mathbb{H}^m$ by

$$v(x^1, \dots, x^m) = \gamma(f(x^1, \dots, x^m)).$$

By the definition of f, we see that v maps the boundary of Ω_0 to the origin 0 in $\mathbb{H}^m$. Let us write $v = (v^1, v^2, v^3, \dots, v^m)$ as $(v^1 + \sqrt{-1}v^2, v^3, \dots, v^m)$. Suppose we have already defined v on Ω_k for any $0 \leq k < n-1$, then, as before, for any $(x^1, x^2, \dots, x^m)$ in Ω_{k+1}, we set:

$$v(x^1, x^2, x^3, \dots, x^m) = (e^{2\theta_1\sqrt{-1}(k+1/2)}(v^1 - \sqrt{-1}v^2)(\hat{x}), v^3(\hat{x}), \dots, v^m(\hat{x})),$$

where $\hat{x} = (e^{2\sqrt{-1}\theta_{k+1}}(x^1 - \sqrt{-1}x^2), x^3, \dots, x^m)$ which is in $W_k \times [0,\infty)^{m-2}$. Thus, we have defined v on $\mathbb{R}^2 \times [0,\infty)^{m-2}$. Now, we can define $g : \mathbb{R}^m \to \mathbb{H}^m$ by setting:

$$g(x^1, x^2, x^3, \dots, x^m) = (v^1(\widetilde{x}), v^2(\widetilde{x}), \epsilon_3 v^3(\widetilde{x}), \dots, \epsilon_m v^m(\widetilde{x})),$$

for any $x \in \mathbb{R}^m$, where $\widetilde{x} = (x^1, x^2, |x^3|, \dots, |x^m|)$, and $\epsilon_i = \mathrm{sign}(x^i)$, $3 \leq i \leq m$, see Figure 5.

Lemma 3.1. *g is Lipschitz on $\mathbb{R}^m$, and is harmonic on the set* $\arg(x^1 + \sqrt{-1}x^2) \neq \theta_k$, $0 \leq k \leq n-1$, $\tau^3 \cdots \tau^m \neq 0$. *Moreover, if we write*

$$g = (g^1, g^2, g^3, \dots, g^m) = (g^1 + \sqrt{-1}g^2, g^3, \dots, g^m)$$

then:

(i)

$$g(x^1, x^2, \dots, x^m) = (e^{2\sqrt{-1}(k+\frac{1}{2})\theta_1}(g^1 - \sqrt{-1}g^2)(\hat{x}), g^3(\hat{x}), \dots, g^m(\hat{x}))$$

where $\hat{x} = (e^{2\sqrt{-1}\theta_{k+1}}(x^1 - \sqrt{-1}x^2), x^3 \dots, x^m)$;

(ii) *for* $i \geq 3$

$$g^i(x^1, x^2, x^3, \dots, -x^i, \dots, x^m) = -g^i(x^1, x^2, x^3, \dots, x^i, \dots, x^m);$$

and

(iii) *if* $j \neq i$ *with* $i \geq 3$, *then*

$$g^j(x^1, x^2, x^3, \dots, -x^i, \dots, x^m) = g^j(x^1, x^2, x^3, \dots, x^i, \dots, x^m).$$

Proof. The first statement of the lemma follows immediately from the definition of g, the fact that f is harmonic and that γ is a geodesic. The proof of (i) is similar to the proof of Lemma 2.1(i). (ii) and (iii) also follow immediately from the definition of g.

Theorem 3.2. *Let g be the map as above. Then there exists a harmonic map $u : \mathbb{R}^m \mapsto \mathbb{H}^m$ such that*

$$\sup_{x\in\mathbb{R}^m} d(u(x), g(x)) < \infty.$$

Moreover, u is nontrivial in the sense that:

(i) *The image of u is not contained in any totally geodesic submanifold of dimension $m-1$ in of $\mathbb{H}^m$; and*
(ii) *u cannot be decomposed as $u = F \circ G$, such that F is an isometry of $\mathbb{R}^m$, and $G = G(y^1, \ldots, y^{m-1})$ which is independent of the last coordinate.*

Proof. Let B_R be the ball of radius R in $\mathbb{R}^m$ with center at the origin, and let u_R be the harmonic map from B_R to $\mathbb{H}^m$ with $u_R = g$ on ∂B_R. By Lemma 3.1, and the uniqueness of harmonic maps, if we write

$$u_R = (u_R^1, u_R^2, u_R^3, \ldots, u_R^m) = (u_R^1 + \sqrt{-1}u_R^2, u_R^3, \ldots, u_R^m)$$

then

(3.1)
$$u_R(x^1, x^2, \ldots, x^m) = \left(e^{2\sqrt{-1}(k+\frac{1}{2})\theta_1}(u_R^1 - \sqrt{-1}u_R^2)(\hat{x}), u_R^3(\hat{x}), \ldots, u_R^m(\hat{x})\right)$$

where $\hat{x} = (e^{2\sqrt{-1}\theta_{k+1}}(x^1 - \sqrt{-1}x^2), x^3 \ldots, x^m)$; for $i \geq 3$

$$u_R^i(x^1, x^2, x^3, \ldots, -x^i, \ldots, x^m) = -u_R^i(x^1, x^2, x^3, \ldots, x^i, \ldots, x^m), \tag{3.2}$$

and if $j \neq i$, with $i \geq 3$,

$$u_R^j(x^1, x^2, x^3, \ldots, -x^i, \ldots, x^m) = u_R^j(x^1, x^2, x^3, \ldots, x^i, \ldots, x^m). \tag{3.3}$$

We want to prove that there is a constant C which is independent of R such that

$$\sup_{x \in B_R \cap \Omega_0} d(u_R(x), g(x)) \leq C. \tag{3.4}$$

Note that $\partial(B_R \cap \Omega_0) = (\partial B_R \cap \Omega_0) \cup (\partial\Omega_0 \cap B_R)$. On $\partial B_R \cap \Omega_0$, $u_R = g$. $\partial\Omega_0 \cap B_R$ consists of those points $(x^1, x^2, \ldots, x^m) \in B_R$ such that $\arg(x^1 + \sqrt{-1}x^2) = \theta_0$ or θ_1. By (3.1), if $\arg(x^1 + \sqrt{-1}x^2) = 0$, then as in the proof of Theorem 2.2, we have $\arg(u_R^1(x) + \sqrt{-1}u_R^2(x)) = -\frac{1}{2}\theta_1$ or $\pi - \frac{1}{2}\theta_1$. By the definition of g, it is easy to see that if $(a^1, a^2, \ldots, a^m) \in \overline{g(\mathbb{R}^m)} \cap \partial\mathbb{H}^m$, then there exists k, such that

$$a^1 + \sqrt{-1}a^2 = \frac{e^{i\theta_k}}{\sqrt{m-1}} \neq 0.$$

As in the proof of Theorem 2.2, we conclude that

$$\Pi \cap \overline{g(\mathbb{R}^m)} \cap \partial\mathbb{H}^m = \emptyset$$

where Π is the hyperplane $(v^1, v,^2, \ldots, v^m)$, such that $\arg(v^1 + \sqrt{-1}v^2) = -\frac{1}{2}\theta_1$ or $\pi - \frac{1}{2}\theta_1$. Hence there is a constant C_1 which is independent of R such that

$$d(u_R(x), 0) \leq C_1$$

for all $x \in \partial\Omega_0 \cap B_R$ with $\arg(x^1 + \sqrt{-1}x^2) = \theta_0 = 0$. Note that for such x, $g(x) = 0$. Hence

$$d(u_R(x), g(x)) \leq C_1 \tag{3.5}$$

for all $x \in \partial\Omega_0 \cap B_R$ with $\arg(x^1 + \sqrt{-1}x^2) = \theta_0 = 0$. Similarly, one can show that (3.5) is true for $x \in \partial\Omega_0 \cap B_R$ with $\arg(x^1 + \sqrt{-1}x^2) = \theta_1$. By the maximum principle, we conclude that (3.4) is true. By Lemma 3.1 of g and (3.1)–(3.3), we see that

$$\sup_{x \in \mathbb{R}^m} d(u_R(x), g(x)) \leq C_1$$

for some constant C_1 which is independent of R. Passing to a subsequence if necessary, we can find a harmonic map u from $\mathbb{R}^m$ into $\mathbb{H}^m$ such that

$$\sup_{x \in \mathbb{R}^m} d(u(x), g(x)) \leq C_1. \tag{3.6}$$

From this we have

$$\overline{u(\mathbb{R}^m)} \cap \partial\mathbb{H}^m = \overline{g(\mathbb{R}^m)} \cap \partial\mathbb{H}^m.$$

The set on the right hand side contains all points of the form

$$\frac{1}{\sqrt{m-1}}(\cos\theta_k, \sin\theta_k, a^3, \ldots, a^m)$$

for some k, where a^j is either $+1$ or -1, for $3 \leq j \leq m$. Hence the set cannot be contained in any hyperplane in $\mathbb{R}^m$. We conclude that $u(\mathbb{R}^m)$ is not contained in any totally geodesic submanifold of dimension $m-1$ in of $\mathbb{H}^m$. This proves (i). To prove (ii), we may assume that F is a linear isomorphism and it is sufficient to show that for any $(m-1)$ dimensional subspace $\mathbb{P}$ of $\mathbb{R}^m$, $\overline{u(\mathbb{P})} \cap \partial\mathbb{H}^m \neq \overline{u(\mathbb{R}^m)} \cap \partial\mathbb{H}^m$. By (3.6), it is sufficient to show that

$$\overline{g(\mathbb{P})} \cap \partial\mathbb{H}^m \neq \overline{g(\mathbb{R}^m)} \cap \partial\mathbb{H}^m. \tag{3.7}$$

Since $\mathbb{P}$ is a proper subspace, there is some fixed ϵ_i which is either $+1$ or-1, $3 \leq i \leq m$, and there is some k such that if

$$\Omega = \{(x^1, \ldots, x^m) |\ (x^1, x^2, \epsilon_3 x^3, \ldots, \epsilon_3 x^m) \in \Omega_k\}$$

then $\mathbb{P}$ will not intersect the interior of Ω. By the definition of g, we see that (3.7) is true.

Again the harmonic map u in the theorem is of polynomial growth, and the closure of its image intersects the geometric boundary at $n \times 2^{m-2}$ points.

4. Asymptotic behaviors of harmonic diffeomorphisms from $\mathbb{R}^2$ into $\mathbb{H}^2$.

In this section, we will discuss the asymptotic behavior of a harmonic diffeomorphism u from $\mathbb{R}^2$ into $\mathbb{H}^2$ with Hopf differential ϕdz^2 such that ϕ is a polynomial. It was proved in [**HTTW**] that the image of such a map is an ideal polygon. However, it is unclear how to determine the exact positions of the vertices of the polygon in terms of ϕ. On the other hand, it is also proved that each horizontal ray of ϕdz^2 is mapped under u into a curve which is asymptotically a geodesic ray in $\mathbb{H}^2$. In this section we want to show that the behavior of a harmonic diffeomorphism with Hopf differential $z^n dz^2$ is rather typical, in the sense that the image of the harmonic map along each ray in certain directions will tend to infinity at a rate depending only on n and the direction of the ray. While the results may have interest in their own right, they will be applied in the next section to study the relation between the Hopf differential and the image of a harmonic map from $\mathbb{R}^2$ to $\mathbb{H}^2$.

To fix notations, let u be an orientation preserving harmonic diffeomorphism from $\mathbb{R}^2$ into $\mathbb{H}^2$, so that its Hopf differential is of the form ϕdz^2 where

$$\phi(z) = z^n + \sum_{j=1}^{n} a_j z^{n-j} = z^n(1+h)$$

and

$$h(z) = \sum_{j=1}^{n} a_j z^{-j}.$$

Lemma 4.1. *Let θ be such that $\cos((\frac{n}{2}+1)\theta) \neq 0$, and let $L(T,\theta)$ be the length of $u(te^{i\theta})$, $0 \le t \le T$. We have:*

(a) *If $n = 2m$, then as $T \to \infty$*

$$\begin{aligned}\frac{1}{2}L(T,\theta) = \Bigg| &\frac{T^{m+1}}{m+1}\cos\left(\left(\frac{n}{2}+1\right)\theta\right) \\ &+ \sum_{j=1}^{m} \frac{T^{m-j+1}}{m-j+1}\binom{\frac{1}{2}}{j}\Re\left(e^{\sqrt{-1}(m+1-j)\theta}c_j\right) \\ &+ \binom{\frac{1}{2}}{m+1}\log T \cdot \Re(c_{m+1})\Bigg| + O(1).\end{aligned}$$

(b) *If $n = 2m+1$, then as $T \to \infty$*

$$\begin{aligned}\frac{1}{2}L(T,\theta) = \Bigg| &\frac{T^{m+\frac{3}{2}}}{m+\frac{3}{2}}\cos\left(\left(\frac{n}{2}+1\right)\theta\right) \\ &+ \sum_{j=1}^{m+1} \frac{T^{m-j+\frac{3}{2}}}{m-j+\frac{3}{2}}\binom{\frac{1}{2}}{j}\Re\left(e^{\sqrt{-1}(m-j+\frac{3}{2})\theta}c_j\right)\Bigg| + O(1).\end{aligned}$$

Here for each $1 \le j \le n$, c_j are functions of $a_1, \ldots, a_j$. $\Re(z)$ is the real part of the complex number z.

Proof. Since there exists $R > 0$ such that $\phi(z) \neq 0$ outside $|z| < R$, by deleting a half line, we can choose a branch of $\sqrt{\phi}$ on $|z| > R$. We may assume that $te^{\sqrt{-1}\theta}$ is not on the deleted half line. Let $\xi + i\eta = w =$

$\int^z \sqrt{\phi(\zeta)}d\zeta$. Then locally, w is a complex coordinates of $\mathbb{R}^2$. The pull-back metric of $\mathbb{H}^2$ under u is

$$(e+2)d\xi^2 + (e-2)d\eta^2,$$

where e is energy density of u with respect to the metric $|dw|^2 = |\phi||dz|^2$. Let $z = te^{i\theta}$, then

$$\frac{dw}{dt} = \frac{dw}{dz}\frac{dz}{dt} = e^{i\theta}\sqrt{\phi(te^{\sqrt{-1}\theta})}. \tag{4.1}$$

By [**Hn**], there is a constant $C_1 > 0$ such that

$$0 \le e(z) - 2 \le \exp(-C_1|z|). \tag{4.2}$$

Also $|h| < 1$ when $|z|$ is large,

$$\begin{aligned} \left|\frac{d\xi}{dt}\right| &= \left|\Re\left\{e^{\sqrt{-1}\theta}\sqrt{\phi(te^{\sqrt{-1}\theta})}\right\}\right| \\ &= \left|\Re\left\{t^{\frac{n}{2}}e^{\sqrt{-1}(\frac{n}{2}+1)\theta}\left(1+\sum_{j=1}^{\infty}\binom{\frac{1}{2}}{j}h^j\right)\right\}\right| \\ &= t^{\frac{n}{2}}\left|\left(\cos\left(\left(\frac{n}{2}+1\right)\theta\right)+\sum_{j=1}^{n}t^{-j}\binom{\frac{1}{2}}{j}\Re\left(e^{\sqrt{-1}(\frac{n}{2}+1-j)\theta}c_j\right)\right.\right. \\ &\qquad\left.\left. + O(t^{-n-1})\right)\right| \end{aligned} \tag{4.3}$$

as $t \to \infty$. Since $\cos((\frac{n}{2}+1)\theta) \ne 0$, we have

$$\begin{aligned} (e+2)\left|\frac{d\xi}{dt}\right|^2 + (e-2)\left|\frac{d\eta}{dt}\right|^2 &= 4\left|\frac{d\xi}{dt}\right|^2 + (e-2)\left|\frac{dw}{dt}\right|^2 \\ &= 4\left|\frac{d\xi}{dt}\right|^2 + O(\exp(-C_2 t)). \end{aligned}$$

Hence there exists $t_0 > 0$ such that if $t > t_0$,

$$\begin{aligned} L(t,\theta) &= \left|\int_{t_0}^{t}\sqrt{(e+2)\left|\frac{d\xi}{dt}\right|^2 + (e-2)\left|\frac{d\eta}{dt}\right|^2}\right| + O(1) \\ &= \int_{t_0}^{t} 2\left|\frac{d\xi}{dt}\right| + O(1). \end{aligned}$$

Using (4.3), the lemma follows.

Remark 4.1. As one can see from the proof, even if $\cos((\frac{n}{2}+1)\theta) = 0$, we still have $\lim_{t\to\infty} L(t,\theta) = \infty$, provided that one of the coefficients of the term T^{m-j+1} or $\log T$ is not zero for the case $n = 2m$. The situation for $n = 2m+1$ is similar.

Proposition 4.2. *With the notations and assumptions as in Lemma* 4.1, *we have*

$$\lim_{t\to\infty}\frac{d(o,u(te^{\sqrt{-1}\theta}))}{L(t,\theta)} = \lim_{t\to\infty}\frac{(\frac{n}{2}+1)d(o,u(te^{\sqrt{-1}\theta}))}{t^{\frac{n}{2}+1}\cos((\frac{n}{2}+1)\theta)}$$
$$= 1$$

where o *is a fixed point in* $\mathbb{H}^2$.

Proof. Let $\gamma(t) = u(te^{\sqrt{-1}\theta})$, and let $w = \xi + i\eta$ be as in the proof of Lemma 4.1. In these coordinates, the geodesic curvature of $\gamma(t)$ is

(4.4)
$$\kappa(t) = (\dot{t})^3\sqrt{e^2-4}\Big[\Gamma^2_{11}(\xi')^3 + (2\Gamma^2_{12}-\Gamma^1_{11})(\xi')^2\eta' - (2\Gamma^1_{12}-\Gamma^2_{22})\xi'(\eta')^2 - \Gamma^1_{22}(\eta')^3 + \xi'\eta'' - \xi''\eta'\Big]$$

where

$$\Gamma^1_{11} = \frac{1}{2}(e+2)^{-1}\frac{\partial e}{\partial\xi},\quad \Gamma^1_{12} = \frac{1}{2}(e+2)^{-1}\frac{\partial e}{\partial\eta},\quad \Gamma^1_{22} = \frac{1}{2}(e+2)^{-1}\frac{\partial e}{\partial\xi},$$
$$\Gamma^2_{11} = -\frac{1}{2}(e-2)^{-1}\frac{\partial e}{\partial\eta},\quad \Gamma^2_{12} = \frac{1}{2}(e-2)^{-1}\frac{\partial e}{\partial\xi},\quad \Gamma^2_{22} = \frac{1}{2}(e+2)^{-1}\frac{\partial e}{\partial\eta},$$

$\dot{t} = \frac{dt}{ds}$, s is the arc length of $\gamma(t)$ and e is the energy density of u with respect to the metric $|dw|^2 = |\phi|\,|dz|^2$. As in [**Hn**], we have

$$(e-2)^{-\frac{1}{2}}|\nabla e| \le C_1\exp(-C_2|z|) \tag{4.5}$$

for z large enough, and ∇ is the gradient with respect to the metric $|\phi||dz|^2 = |dw|^2$. Since $\cos((\frac{n}{2}+1)\theta) \neq 0$, by (4.3)

$$\frac{ds}{dt} = \sqrt{(e+2)\left|\frac{d\xi}{dt}\right|^2 + (e-2)\left|\frac{d\eta}{dt}\right|^2} = 2\left|\frac{d\xi}{dt}\right| + O(\exp(-Ct)). \tag{4.6}$$

Note that we also have

$$\left|\frac{dw}{dt}\right| \le C_2 t^{n/2}, \tag{4.7}$$

$$t^{-\frac{n}{2}}\left|\frac{d\xi}{dt}\right| = \left|\cos\left(\left(\frac{n}{2}+1\right)\theta\right)\right| + o(1). \tag{4.8}$$

$$\left|\frac{d^2w}{dt^2}\right| \le C_2 t^{n/2-1} \tag{4.9}$$

for some constants C_2, C_3. By (4.4)–(4.9), we have

$$|\kappa(t)| \le C_4\exp(-C_5 t) \tag{4.10}$$

for some positive constants C_4 and C_5. By (4.10) and Lemma 3.1 in [**HTTW**], given $\epsilon > 0$, there is $t_0 > 0$, and a geodesic line α passing through $\gamma(t_0)$ such that

$$d(\gamma(t), \alpha) \leq \epsilon \tag{4.11}$$

for all $t > t_0$. Let $f^2 d\rho^2 + d\tau^2$ be the Fermi coordinates with respect to α, so that $\tau = 0$ is the geodesic α, where $f = \cosh \tau$. Under this coordinates, $\gamma(t) = (\rho(t), \tau(t))$. By (4.10), we have at $\gamma(t)$

$$|\ddot{\tau} - f f_\tau (\dot{\rho})^2| \leq C_4 \exp(-C_5 t),$$

and so

$$\begin{aligned} |\ddot{\tau}| &\leq C_4 \exp(-C_5 t) + |f f_\tau (\dot{\rho})^2| \\ &\leq C_4 \exp(-C_5 t) + C_6 \epsilon |f(\dot{\rho})^2| \\ &\leq C_4 \exp(-C_5 t) + C_6 \epsilon |f^2(\dot{\rho})^2| \\ &\leq C_7 \epsilon \end{aligned}$$

for some constants C_6, C_7, provided t_0 is large enough, where we have used the fact that $|\tau| \leq \epsilon$, $f = \cosh \tau$ and the fact that $f^2(\dot{\rho})^2 \leq 1$. Here and below, "$\dot{}$" means differentiation with respect to arc length s and where "$'$" means differentiation with respect to t. Hence

$$\left| \frac{d}{dt}(\dot{\tau}) \right| = \left| \ddot{\tau} \frac{ds}{dt} \right| \leq C_7 \epsilon \left| \frac{ds}{dt} \right|. \tag{4.12}$$

Since $|\dot{\gamma}| = 1$, we have

$$f^2(\dot{\rho})^2 + (\dot{\tau})^2 = 1.$$

For any $T > t_0$, suppose $\tau'(T) = 0$, then

$$f^2(\dot{\rho})^2 = 1,$$

at T. Suppose $\tau'(T) \neq 0$, let us we assume $\tau'(T) > 0$, the case that $\tau'(T) < 0$ is similar. Let b be the supremum of c such that $\tau' > 0$ on $[T, T+c)$. Suppose $b < \infty$, then $\tau'(T + b) = 0$. By (4.12),

$$\left| \frac{d}{dt}[(\dot{\tau})^2] \right| = 2|\dot{\tau}| \left| \frac{d}{dt}(\dot{\tau}) \right| \leq C_7 \epsilon |\dot{\tau}| \left| \frac{ds}{dt} \right| = C_7 \epsilon \frac{d\tau}{dt}$$

in $(T, T + b)$. Hence

$$\begin{aligned} (\dot{\tau})^2(T) - (\dot{\tau})^2(T+b) &\leq \int_T^{T+b} \left| \frac{d}{dt}[(\dot{\tau})^2] \right| dt \\ &\leq C_7 \epsilon \int_T^{T+b} \frac{d\tau}{dt} \\ &\leq C_7 \epsilon^2 \end{aligned}$$

where we have used the fact that $|\tau| \le \epsilon$. Since $(\dot\tau)^2(T+b)=0$, we have

$$(\dot\tau)^2(T) \le C_7\epsilon^2$$

and so

$$f^2(\dot\rho)^2 \ge 1 - C_7\epsilon^2$$

at T. If $b=\infty$, then we can choose $t_i \to \infty$ with $\tau'(t_i) \to 0$, and we obtain the same inequality. In particular, $f^2(\dot\rho)^2(T)$ is not 0 for all $T > t_0$, provided t_0 is large enough. Without loss of generality, we may assume that $f\dot\rho > 0$ on $[t_0, \infty)$. For any $T > t_0$,

$$\begin{aligned}\rho(T) - \rho(t_0) &= \int_{t_0}^{T} \frac{d\rho}{dt} dt \\ &= \int_{t_0}^{T} \dot\rho \frac{ds}{dt} dt \\ &= \int_{t_0}^{T} f^{-1} f\dot\rho \frac{ds}{dt} dt \\ &\ge (1 - C_8\epsilon)(s(T) - s(t_0))\end{aligned}$$

for some constant C_8. So

$$d(o, u(Te^{i\theta})) \ge \rho(T) - \tau(T) \ge (1 - C_8\epsilon)(s(T) - s(t_0)) - \epsilon.$$

It is obvious that,

$$d(o, u(Te^{i\theta})) \le s(T).$$

Note that $s(T) = L(T,\theta)$ in our previous notation and the lemma follows easily.

5. Hopf differentials and images of harmonic maps.

In [**HTTW**], it was proved that if u is a harmonic diffeomorphism from $\mathbb{R}^2$ into $\mathbb{H}^2$ with polynomial Hopf differential, then its image is an ideal polygon. In this section, we will use the analysis in §4 to study explicit relation in some special cases between the Hopf differential and the position of the vertices of the image of u.

Theorem 5.1. *Let $\phi(z) = z^{2m} + az^{m-1}$, where a is a real number. Suppose u is an orientation preserving harmonic diffeomorphism from $\mathbb{R}^2$ to $\mathbb{H}^2$ with Hopf differential ϕdz^2. Then by composing an isometry of $\mathbb{H}^2$ if necessary, the image of u is a regular ideal polygon.*

Proof. Let $w = \log ||\partial u||$, where $||\partial u|| = \sigma|\frac{\partial u}{\partial z}|$ and $\sigma^2|du|^2$ is the metric on $\mathbb{H}^2$. Then w is the unique solution of

$$\Delta_0 w = e^{2w} - |\phi|^2 e^{-2w} \tag{5.1}$$

such that $e^{2w}|dz|^2$ is a complete metric on $\mathbb{R}^2$, see [**WA**]. Here Δ_0 is the Laplacian on $\mathbb{R}^2$ with respect to the standard metric $|dz|^2$. Observe that ϕ satisfies

$$\begin{cases} \phi(z) & = \overline{\phi(\bar{z})}, \quad \text{and} \\ \phi(e^{2\sqrt{-1}\theta}\bar{z}) & = e^{4m\sqrt{-1}\theta}\overline{\phi(z)} \end{cases} \tag{5.2}$$

where $\theta = \frac{\pi}{m+1}$. Identify $\mathbb{H}^2$ with the unit disk $\{u|\ |u| < 1\}$ in $\mathbb{C}$ with Poincaré metric $\sigma^2|du|^2$. Without loss of generality, we may assume that $u(0) = 0$. By Proposition 4.2, we know that if t is real, then $d(u(t), 0) \to \infty$, as $t \to \infty$. We may also assume that $u(t_k)$ tends to the point 1 on the boundary of $\mathbb{H}^2$ for some $t_k \to \infty$, with t_k to be real. Let $v(z) = \overline{u(\bar{z})}$. It is easy to see that v is also an orientation preserving harmonic diffeomorphism. Moreover, let $\zeta = \bar{z}$

$$\begin{aligned} \sigma(v(z))\left|\frac{\partial v}{\partial z}\right|(z) &= \sigma(\overline{u(\bar{z})})\left|\frac{\partial \bar{u}}{\partial \bar{\zeta}}\right|(\zeta) \\ &= \sigma(u(\zeta))\left|\frac{\partial u}{\partial \zeta}\right|(\zeta) \\ &= e^{w(\bar{z})}. \end{aligned}$$

Hence if we let

$$\widetilde{w}(z) = \log\left(\sigma(v(z))\left|\frac{\partial v}{\partial z}\right|(z)\right)$$

then $\widetilde{w}(z) = w(\bar{z})$. By (5.2), it is easy to see that $\widetilde{w}(z)$ also satisfies (5.1), such that $e^{2\widetilde{w}}|dz|^2$ is complete. By uniqueness, we have

$$w(\bar{z}) = \widetilde{w}(z) = w(z). \tag{5.3}$$

On the other hand,

$$\begin{aligned} \sigma^2(v(z))\frac{\partial v}{\partial z}(z)\frac{\partial \bar{v}}{\partial z}(z) &= \sigma^2(\overline{u(\bar{z})})\overline{\frac{\partial u(\zeta)}{\partial \zeta}}\frac{\partial u(\zeta)}{\partial \bar{\zeta}} \\ &= \overline{\sigma^2(u(\zeta))\frac{\partial u(\zeta)}{\partial \zeta}\frac{\partial \bar{u}(\zeta)}{\partial \zeta}} \\ &= \overline{\phi(\zeta)} \\ &= \overline{\phi(\bar{z})} \\ &= \phi(z). \end{aligned} \tag{5.4}$$

By (5.3), (5.4) and the result in [**TW**], $v = \iota \circ u$ for some orientation preserving isometry ι of $\mathbb{H}^2$. Note that $v(0) = \overline{u(0)} = 0$, and for real number t, $v(t) = \overline{u(t)}$. Since we have normalized u so that $u(t_k) \to 1$ as $t \to \infty$, we also have $v(t_k) \to 1$. So ι must be the identity map, and $v \equiv u$. That is to say

$$\overline{u(\bar{z})} = u(z). \tag{5.5}$$

In particular, $u(t)$ is real if t is real. Hence there is $0 < c < 1$ such that the set consisting of those real ξ with $c < \xi < 1$ is in the image of the positive real axis under the harmonic map u. Let $v_1(z) = e^{2\sqrt{-1}\theta}\overline{u(e^{2\sqrt{-1}\theta}\bar{z})}$. Using similar method and (5.2), one can show that $v_1(z) = \iota_1 \circ u(z)$ for some isometry of $\mathbb{H}^2$. Since $v_1(0) = 0 = u(0)$, we have $v_1(z) = e^{\sqrt{-1}\alpha}u(z)$ for some real number α. Hence

$$u(e^{2\sqrt{-1}\theta}\bar{z}) = e^{2\sqrt{-1}\beta}\overline{u(z)} \tag{5.6}$$

where $2\beta = 2\theta - \alpha$. We want to prove that $e^{\sqrt{-1}s\beta}$, $0 \leq s \leq 2m+1$ are distinct $(2m+2)$th roots of unity. Moreover the image of u is the ideal polygon spanned by the $e^{\sqrt{-1}s\beta}$, $0 \leq s \leq 2m+1$. This will conclude the proof of the theorem. First, we claim that for any real number t,

$$u(e^{\sqrt{-1}(s+2)\theta}t) = e^{2\sqrt{-1}\beta}u(e^{\sqrt{-1}s\theta}t) \tag{5.7}$$

for all integer $0 \leq s \leq 2m+1$. For $s = 0$, (5.7) follows from (5.6) and (5.5) by letting $z = t$. Suppose (5.7) is true for $0 \leq s < 2m+1$. By (5.6) and (5.5)

$$\begin{aligned} u(e^{\sqrt{-1}(s+3)\theta}t) &= e^{2\sqrt{-1}\beta}\overline{u(e^{-\sqrt{-1}(s+1)\theta}t)} \\ &= e^{2\sqrt{-1}\beta}u(e^{\ i(s+1)\theta}t). \end{aligned}$$

Hence (5.7) is true. By (5.7), we have

$$u(e^{2\sqrt{-1}s\theta}t) = e^{2\sqrt{-1}s\beta}u(t) \tag{5.8}$$

for any integer s. Take $s = m+1$, we have

$$e^{2\sqrt{-1}(m+1)\beta} = 1.$$

By Proposition 4.2, for any $0 \leq s \leq 2m+1$, $d(u(te^{\sqrt{-1}s\theta}), 0) \to \infty$ as $t \to \infty$, t is real. Hence there exists $t_k \to \infty$, and real number b_s such that

$$u(t_k e^{\sqrt{-1}s\theta}) \to e^{\sqrt{-1}b_s},$$

for $0 \leq s \leq 2m+1$. Obviously, $b_0 = 0$, $b_s = \sqrt{-1}s\beta$ for s even by (5.8). On the other hand, by (5.6)

$$\begin{aligned} u(te^{\sqrt{-1}\theta}) &= u(e^{2\sqrt{-1}\theta} \cdot te^{-\sqrt{-1}\theta}) \\ &= e^{2\sqrt{-1}\beta}\overline{u(te^{\sqrt{-1}\theta})}. \end{aligned} \tag{5.9}$$

So we have

$$e^{\sqrt{-1}b_1} = e^{2\sqrt{-1}(\beta - b_1)},$$

and we may assume $b_1 = \beta$ be adding a multiple of 2π to 2β, which does not affect the previous arguments. By (5.7), we again have $b_s = \sqrt{-1}s\beta$ if s is odd. Hence $e^{\sqrt{-1}s\beta}$, $0 \le s \le 2m+1$, are in the closure of the image of u in $\mathbb{H}^2 \cup \partial\mathbb{H}^2$. It remains to prove the $e^{\sqrt{-1}s\beta}$ are distinct. Suppose not, then $e^{\sqrt{-1}s\beta} = 1$ for some $0 < s \le 2m+1$. If s is even, then by (5.8), we have a contradiction, because u is one to one and $e^{\sqrt{-1}s\theta} \ne 1$. Suppose, s is odd. By (5.9), we have $u(te^{\sqrt{-1}\theta}) = \rho(t)e^{\sqrt{-1}\beta}$ where $\rho(t) > 0$. By (5.7),

$$u(te^{\sqrt{-1}s\theta}) = \rho(t)e^{\sqrt{-1}s\beta} = \rho(t).$$

Since $\rho(t_k) \to 1$ and $c < \xi < 1$ is in the image of the positive real axis under u, this contradicts the fact the u is one to one. The theorem follows from the fact that the image of u is a ideal polygon of $2m+2$ sides [**HTTW**].

Next we will discuss the Hopf differentials of the harmonic diffeormorphisms constructed in Proposition 1.6.

Proposition 5.2. *Let $u(z)$ be the harmonic diffeomorphism constructed in Proposition* 1.6. *Then there is a conformal map $z = z(\zeta)$ such that the Hopf differential of u with respect to ζ is of the form $(\zeta^2 + \sqrt{-1}\alpha)d\zeta^2$ where α is a real number.*

Proof. Let $u(z)$ be the harmonic diffeomorphism constructed in Proposition 1.6. Then $u(\bar{z}) = \overline{u(z)}$ and $u(-\bar{z}) = -\overline{u(z)}$. Let $\phi(z)dz^2$ be the Hopf differential of u, then

$$\phi(z) = \sigma^2(u(z))\frac{\partial u}{\partial z}\frac{\partial \bar{u}}{\partial z}.$$

It is easy to see that $\phi(\bar{z}) = \overline{\phi(z)}$ and $\phi(-\bar{z}) = \overline{\phi(z)}$. By the result of [**HTTW**], ϕ is a polynomial of degree 2, that is $\phi(z) = az^2 + bz + c$. Now $\phi(\bar{z}) = \overline{\phi(z)}$ implies that a, b and c are real. $\phi(-\bar{z}) = \overline{\phi(z)}$ implies that $b = 0$. Hence $\phi(z) = az^2 + c$, where a and c are real. Let β be any one of the fourth root of a, and let $\zeta = \beta z$, then

$$\begin{aligned}\phi(z)dz^2 &= (az^2 + c)dz^2 \\ &= (a\beta^{-2}\zeta + c)\beta^{-2}d\zeta^2 \\ &= (\zeta^2 + \sqrt{-1}\alpha)d\zeta^2\end{aligned}$$

where $\sqrt{-1}\alpha = c\beta^{-2}$. Suppose $a > 0$, then we may choose β to be a positive real number. Hence $\sqrt{-1}\alpha$ is real. By Remark 4.1, the length of the image under u of the half line $\zeta > 0$ is infinite. On the other hand, $\beta > 0$, $\zeta > 0$ implies $z = \beta^{-1}\zeta$ is real and positive. However, by the construction of u in Proposition 1.6, the image of $z > 0$ under u has finite length. This is a contradiction. So $a < 0$, and we may choose $\beta = |a|^{\frac{1}{4}}e^{\frac{\pi}{4}\sqrt{-1}}$. Then

$$\sqrt{-1}\alpha = c\beta^{-2} = c|a|^{\frac{1}{2}}e^{\frac{\pi}{2}\sqrt{-1}} = \sqrt{-1}c|a|^{\frac{1}{2}}.$$

This implies that α is real, because c is real.

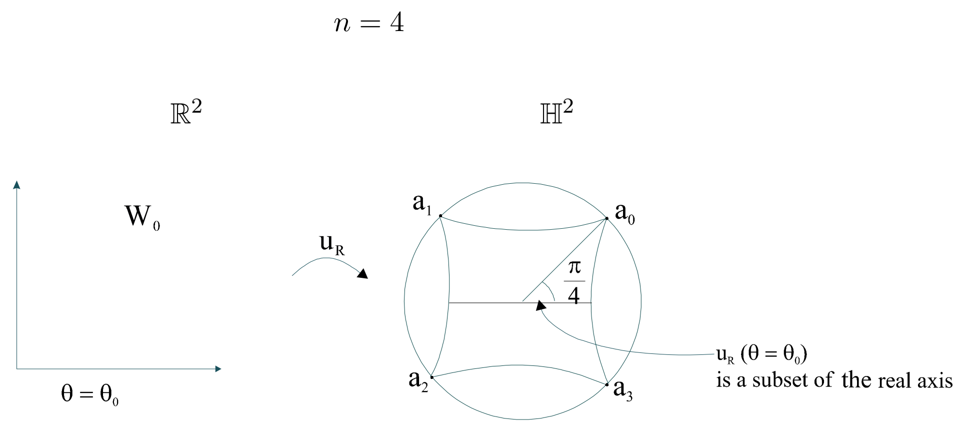

Figure 1

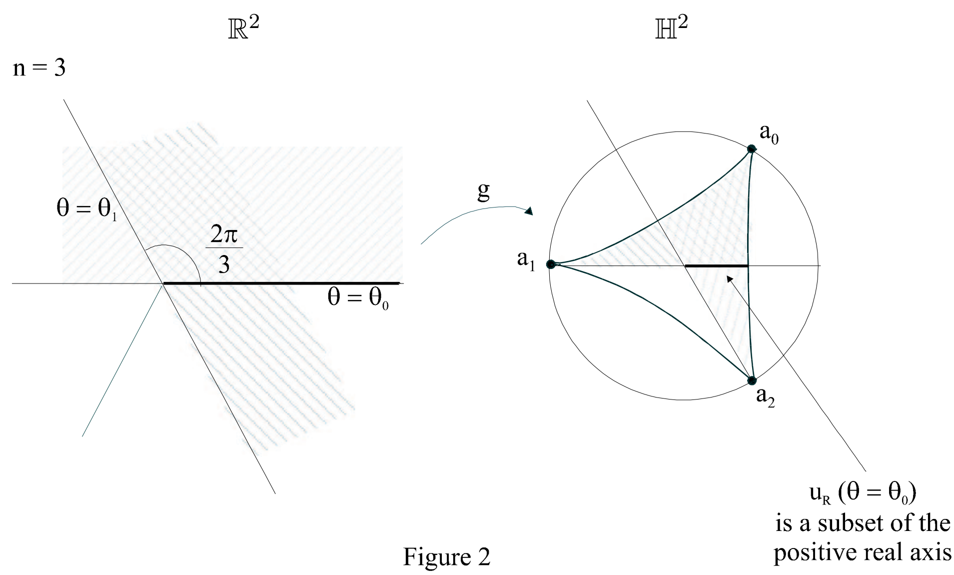

Figure 2

$n = 6$

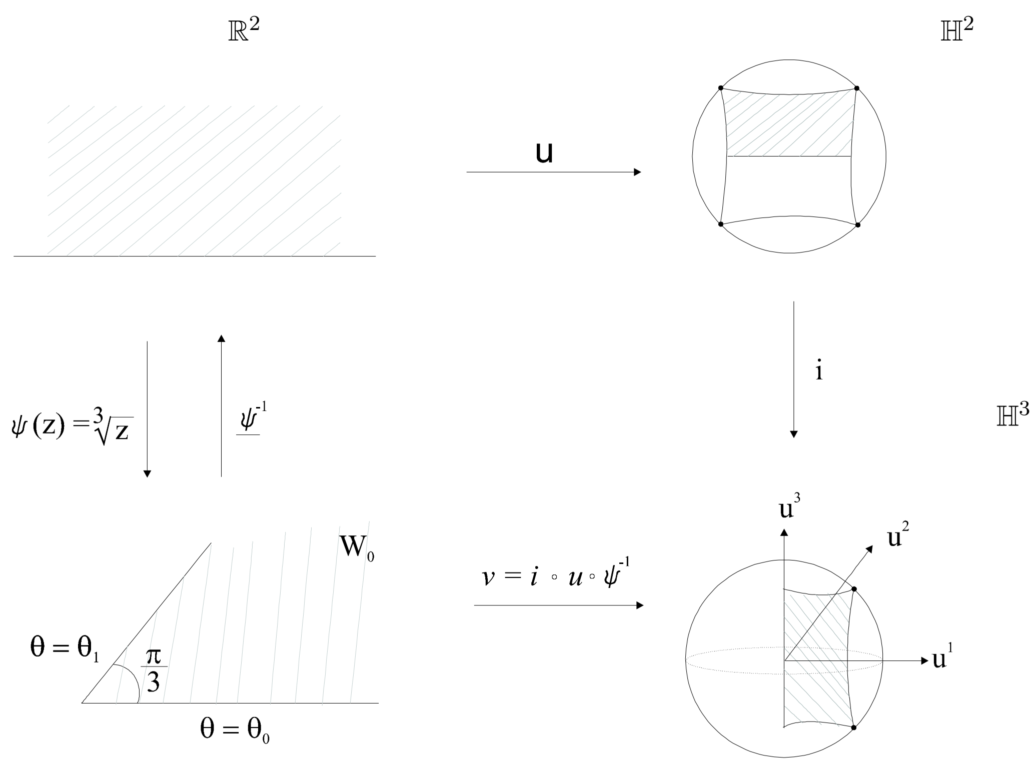

Figure 3

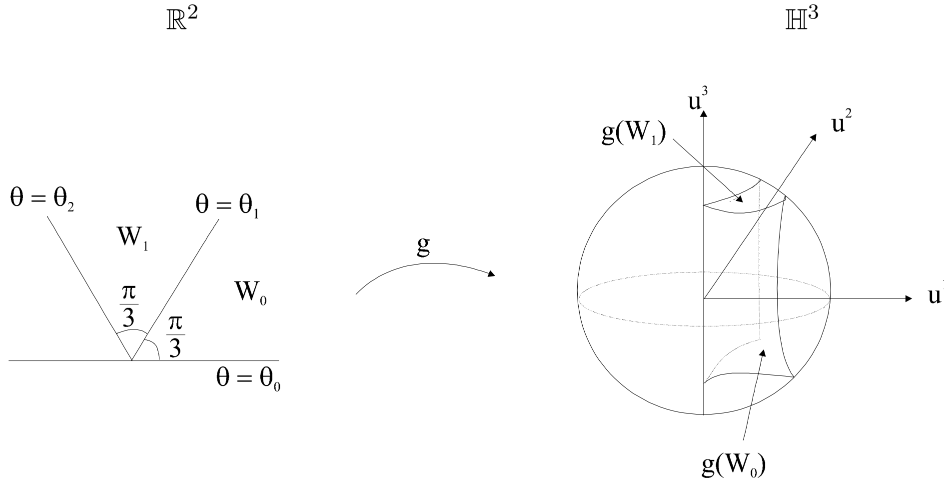

Figure 4

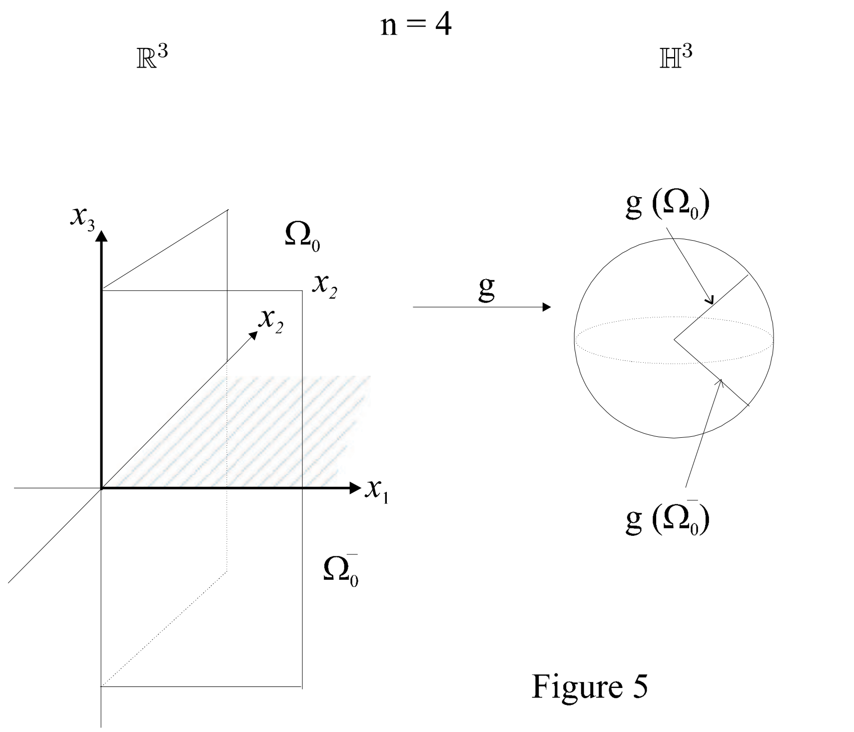

Figure 5

References

[Cg] S.Y. Cheng, *Liouville theorem for harmonic maps*, Proc. Symp. Pure Math., **36** (1980), 147-151, MR 81i:58021, Zbl 455.58009.

[CT] H.I. Choi and A. Treibergs, *Gauss map of spacelike constant mean curvature hypersurface of Minkowski Space*, J. Diff. Geom., **32** (1990), 775-817, MR 92e:58051, Zbl 717.53038.

[Hn] Z.C. Han, *Some remarks concerning the geometric behavior of harmonic maps between surfaces*, preprint, 1994.

[HTTW] Z. Han, L.F. Tam, A. Treibergs and T. Wan, *Harmonic maps from the complex plane into surfaces with nonpositive curvature*, Comm. Anal. Geom., **3** (1995), 85-114, MR 96k:58057, Zbl 843.58028.

[Hz] E. Heinz, *Über die Losungen der Minimalflächengleichung*, Nach. Akad. Wiss. Göttingen Math. Phys. Kl., **II** (1952), 51-56, MR 14,885c.

[J] J. Jost, *Harmonic maps between surfaces*, Lecture Notes in Math., **1062**, Springer-Verlag, 1984, MR 85j:58046, Zbl 542.58002.

[LW] P. Li and J.-P. Wang, *Convex hull properties of harmonic maps*, J. Diff. Geom., **48**(**3**) (1998), 497-530, MR 99f:58050, Zbl 926.58011.

[My] Y. Minsky, *Harmonic maps, length, and energy in Teichmüller space*, J. Diff. Geom., **35** (1992), 151-217, MR 93e:58041, Zbl 763.53042.

[S] R. Schoen, *The role of harmonic mappings in rigidity and deformation problems,* in 'Collection: Complex Geometry (Osaka, 1990)', Lecture Notes in Pure and Applied Mathematics, Dekker, New York, **143** (1993), 179-200, MR 94g:58055, Zbl 806.58013.

[SY] R. Schoen and S.-T. Yau, *On univalent harmonic maps between surfaces*, Invent. Math., **44** (1978), 265-278, MR 57 #17706, Zbl 388.58005.

[Sk] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, **4** (2nd ed.), 1979, Publish or Perish, Inc., MR 82g:53003d, Zbl 439.53004.

[T] A. Tachikawa, *A nonexistence result for harmonic mappings from* $\mathbb{R}^n$ *into* $\mathbb{H}^n$, Tokyo J. Math., **11** (1988), 311-316, MR 89m:58051, Zbl 686.58013.

[TW] L.F. Tam and T.Y.-H. Wan, *Harmonic diffeomorphisms into Cartan-Hadamard surfaces with prescribed Hopf differentials*, Comm. Anal. Geom., **2** (1995), 593-625, MR 96m:58057, Zbl 842.58014.

[WA] T.Y.-H. Wan and T. Au, *Parabolic constant mean curvature spacelike surfaces*, Proc. Amer. Math. Soc., **120** (1994), 559-564, MR 94d:53017, Zbl 798.53059.

[Wf] M. Wolf, *High energy degeneration of harmonic maps between surfaces and rays in Teichmüller space*, Topology, **30** (1991), 517-540, MR 92j:32075, Zbl 747.58027.

Received April 23, 1998 and revised January 20, 1999. The first author's research was supported by NSF Postdoctoral Research Fellowship, China, and MCSEC. The second author's research was partially supported by Earmarked Grant of Hong Kong CUHK325/96P.

Department of Mathematics
Peking University
Beijing, 100871, China

E-mail address: ygshi@math.pku.edu.cn

Department of Mathematics
The Chinese University of Hong Kong
Shatin, NT
Hong Kong, China
E-mail address: lftam@math.cuhk.edu.hk

Guidelines for Authors

Authors may submit manuscripts at pjm.math.berkeley.edu/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

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 will 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, which should be used sparingly.

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.

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.

PACIFIC JOURNAL OF MATHEMATICS

Volume 202 No. 1 January 2002

0030-8730(200201)202:1;1-Y