

Modular curves, SHIMURA curves, and GOPPA codes, better than VARSHAMOV-GILBERT bound

By M. A. TSFASMAN and S. G. VLADUT of Moscow and TH. ZINK of Berlin

(Received September 7, 1981)

Introduction. A linear algebraic code is a k -dimensional linear subspace $C \subset \mathbb{F}_q^n$. The ratio $R = k/n$ is called its transmission rate. A code corrects errors when its elements are “difficult to confuse”, i.e. any two elements $a, b \in C$ have at least d different coordinates (or, which is the same, each vector $a \in C$ has at least d nonzero coordinates). By δ we denote relative minimal distance: $\delta = d/n$.

One of the main problems in algebraic codes is to construct infinite series of codes over fixed \mathbb{F}_q , with n tending to infinity and with R and δ as large as possible. It is known that there exist VARSHAMOV-GILBERT codes with

$$1 - R \approx q(\delta) .$$

where

$$q(x) = x \log_q (q-1) - x \log_q x - (1-x) \log_q (1-x)$$

(\approx meaning asymptotical equality when $n \rightarrow \infty$).

More than twenty years of research made it plausible to think that this boundary is the best possible (cf. [7], 3.3).

V. D. GOPPA has recently discovered some wonderful connections between coding theory and algebraic geometry. He introduced a broad class of codes arising from algebraic curves over finite fields (GOPPA codes, see § 3). Here both the basis in \mathbb{F}_q^n and minimal distance get lucid algebrogeometric interpretation. Good codes are obtained whenever the curve has many \mathbb{F}_q -points in comparison to its genus.

Therefore a natural question arises: having fixed \mathbb{F}_q , how to construct smooth irreducible curves over it such that the ratio of the number of \mathbb{F}_q -points to the genus is large enough? The situation is inverse to the classical one when the genus is fixed and the field varies.

The first two paragraphs of the paper give an answer to this question. In § 1 we study moduli varieties of elliptic curves $X_0(l)$; for their reductions good ratio is obtained over \mathbb{F}_{p^2} . In § 2 we study SHIMURA curves associated with quaternion algebras over real quadratic fields, they have good reductions mod p , forms of these curves have “many” points over \mathbb{F}_p . In § 3 GOPPA codes are introduced and the parameters of those arising from our series of curves are studied.

§ 1 is written by S. G. VLĂDUȚ, § 2 by TH. ZINK, § 3 and introduction by M. A. TSFASMAN.

Let us now put several natural questions we cannot answer:

A. It is well known (A. WEIL) that the number n of \mathbb{F}_q -points of a curve of genus g cannot exceed $1 + q + 2g\sqrt{q}$, i.e. the ratio $\gamma = g/n \cong (2\sqrt{q})^{-1}$ asymptotically with q fixed and g tending to infinity. Coding theory shows (a consequence of PLOTKIN bound) that when $q = 2$ or 3 this value cannot be reached. How small can γ be? Is not the estimate $\gamma \approx (\sqrt{q} - 1)^{-1}$ of paragraphs 1 and 2 the best possible? In other words, how can the eigenvalues of the FROBENIUS acting on the first cohomology group of a curve over \mathbb{F}_q be distributed when q is fixed and the genus tends to infinity?

B. Curves $X_0(l)$ can be given by explicit equations of degree $l + 1$ in \mathbb{P}^2 , but these equations are highly singular. How to describe (for example in terms of modular functions) GOPPA codes arising from these curves? Can such a description be made "constructive"?

C. Moduli varieties studied in §§ 1 and 2 are far from being only known. What can one say about the points on other moduli varieties?

The work is based on excellent ideas of V. D. GOPPA to whom the authors are sincerely grateful. Authors are also deeply grateful to JU. I. MANIN without whom the paper would never have been written.

When this paper was already written, JU. I. MANIN kindly pointed to us that results similar to those of § 1 are contained in IHARA papers [4] v. 2 ch. 5 § 26 and [5]. Namely, it is shown that for the number n of \mathbb{F}_{p^2} -rational points on reductions of modular curves with projective structure of level m , $p \nmid m$, holds $n \cong \cong (g - 1)(p - 1)$. Note that the proof of the theorem of § 1 shows that for any $H \subset GL_2(\mathbb{Z}/m)$ containing the diagonal subgroup we have $n \cong (p - 1)(GL_2(\mathbb{Z}/m) : H)/12$, where n is the number of \mathbb{F}_{p^2} -points of the reduction of modular curve with structure of level H .

IHARA has shown in [6] that there exist families of SHIMURA curves with the same ratio γ we have found in § 2. Our method to construct those families is totally different and has the advantage to single out explicit examples.

§ 1. Here we show that the reductions mod p of modular curves have "many" points defined over \mathbb{F}_{p^2} .

Let $l \neq p$ be a rational prime, $\Gamma_0(l)$ denotes, as usual, the following subgroup of $SL_2(\mathbb{Z})$:

$$\Gamma_0(l) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid c \equiv 0 \pmod{l} \right\}.$$

$\Gamma_0(l)$ acts on the halfplane $H = \{z \in \mathbb{C} \mid \text{Im } z > 0\}$ and defines an analytic space $H/\Gamma_0(l)$. The latter is the set of complex points of an affine algebraic curve $Y_0(l)$ defined over \mathbb{Q} . The study of its reductions is based on the representation of $Y_0(l)$ as moduli variety of elliptic curves with an additional structure. Let us resume some facts we need from [2].

There exists a smooth proper scheme $M_{Y_0(l)}^0[1/l]$ over $\mathbb{Z}[1/l]$ of relative dimension 1 such that for any geometric point $\bar{s} : \text{Spec } \bar{k} \rightarrow \text{Spec } \mathbb{Z}[1/l]$ there is a bijection:

$$M_{Y_0(l)}^0[1/l](\bar{s}) \simeq \left\{ \begin{array}{l} \text{set of isomorphism classes of pairs } (E, \alpha), \\ \text{where } E \text{ is an elliptic curve and } \alpha : E \rightarrow E_1 \\ \text{an isogeny of degree } l. \end{array} \right\}$$

The fibre of $M_{Y_0(l)}^0[1/l]$ at the point $\text{Spec } \mathbb{Q} \rightarrow \text{Spec } \mathbb{Z}[1/l]$ is $Y_0(l)$.

There exists a projection

$$M_{Y_0(l)}^0[1/l] \rightarrow M_{Y_0(1)}^0[1/l] = \mathbb{A}_{\mathbb{Z}[1/l]}^1$$

given by the classical j -invariant.

Denote $Y_0(l)/p = M_{Y_0(l)}^0[1/l] \otimes \mathbb{F}_p$, it is the reduction of $Y_0(l)$ into characteristic p . The projection $j : Y_0(l)/p \rightarrow \mathbb{A}_{\mathbb{F}_p}^1$ is of degree $l+1$.

We shall show that the fibre over a supersingular value of j contains many \mathbb{F}_p -points. Let us denote the smooth compactification of $Y_0(l)/p$ by $X_0(l)/p$ (i.e. $M_{Y_0(l)}^0[1/l] \otimes \mathbb{F}_p$ in the notation of [1]). $X_0(l)/p$ is a smooth and complete curve of genus $[l/12]$.

Theorem. *The number $n = X_0(l)/p(\mathbb{F}_{p^2})$ of \mathbb{F}_{p^2} -rational points on $X_0(l)/p$ satisfies $n \equiv (p-1)(l+1)/12$. Thus $g/n \equiv (p-1)^{-1}$.*

Proof. Remember that $j(E) \in \mathbb{F}_{p^2}$ for all supersingular curves in characteristic p . We know by [2]:

$$(1) \quad \sum_{j \text{ supersing.}} |\text{Aut } E|^{-1} = (p-1)/24.$$

Let us first count the $\bar{\mathbb{F}}_p$ -rational points of $Y_0(l)/p$. If j is supersingular but not 0 or 1728 there are $(l+1)$ $\bar{\mathbb{F}}_p$ -rational points over j . Indeed, because $\text{Aut } E = \{\pm 1\}$ the $(l+1)$ subgroups of order l of $X(\bar{\mathbb{F}}_p)$ give rise to nonisomorphic isogenies. In the same manner we get for any value of j at least $2(l+1)/|\text{Aut } E|^{-1}$ $\bar{\mathbb{F}}_p$ -rational points over j . We conclude by (1) that there are at least $(p-1)(l+1)/12$ $\bar{\mathbb{F}}_p$ -rational points on $Y_0(l)/p$ that are supersingular.

It remains to show that these points are rational over \mathbb{F}_{p^2} . Assume for a moment that for any supersingular j there is an elliptic curve E having this invariant and such that the Frobenius π_E over \mathbb{F}_{p^2} acts by multiplication of p . Then any isogeny $E \rightarrow E_1$ is also defined over \mathbb{F}_{p^2} and hence defines a \mathbb{F}_{p^2} -rational point of $Y_0(l)/p$.

To prove our assumption, let E' be any elliptic curve defined over \mathbb{F}_{p^2} having invariant j . It is well known that

$$\pi_{E'} = p\varepsilon,$$

where ε is an automorphism of E' . Since ε is of finite order we find by descent a form E of E' over \mathbb{F}_{p^2} such that $\pi_E = p$. This completes the proof of the theorem.

Remark. One can show that we have counted almost all \mathbb{F}_{p^2} -rational points on $X_0(l)/p$. To be precise, for any p there exists a constant c_p , such that

$$|X_0(l)/p(\mathbb{F}_{p^2})| \leq (p-1)(l+1)/12 + c_p.$$

§ 2. In the first paragraph we have considered SHIMURA curves of the form H/Γ where Γ was a congruence subgroup of $SL_2(\mathbb{Q})$. We consider now the case where Γ is associated to a quaternion algebra B over a real quadratic number field K . We get complete algebraic curves H/Γ . Our main result in this paragraph is a formula for the genus of H/Γ .

We first state our result. Let p be a rational prime that remains prime in K . Let B be a quaternion algebra over K that is unramified in p and such that

$$B \otimes_{\mathbb{Q}} \mathbb{R} = M_2(\mathbb{R}) \times \mathbb{H}.$$

Let \bar{C} be an open and compact subgroup of $B^*(\mathbb{A}_f)$ of the form $\bar{C} = C^p C_p$, where $C^p \subset B^*(\mathbb{A}_f^p)$ and $C_p \subset B^*(\mathbb{Q}_p)$ is a maximal compact subgroup. We will always assume that \bar{C} is contained in some of the congruence subgroups

$$\bar{C}_n = \{g \in B^*(\mathbb{A}_f) \mid (g-1)(0_B \otimes \hat{\mathbb{Z}}) \subset n(0_B \otimes \hat{\mathbb{Z}})\},$$

where 0_B is some order of B and $n \geq 3$ some rational integer.

Let $M_{B^*, \bar{C}} = \bar{C} \times_{\infty} B^*(\mathbb{A})/B$ be the SHIMURA variety associated to B and \bar{C} (DELIGNE [3] § 6). It is a smooth curve canonically defined over K . It has good reduction modulo p . That means that there exists a smooth projective scheme

$$\mathfrak{M}_{B^*, \bar{C}} \rightarrow \text{Spec } \mathbb{Z}_{(p)}$$

which has $\mathfrak{M}_{B^*, \bar{C}}$ as general fibre. This and all unproven facts that follow the reader will find in ZINK [11]. By DELIGNE [3] the reduced norm induces the STEIN factorization

$$(2) \quad \mathfrak{M}_{B^*, \bar{C}} \rightarrow \pi_0(\mathfrak{M}_{B^*, \bar{C}}) = Nm\bar{C} \backslash K^*(\mathbb{A}_f)/K_+^*.$$

K_+^* denotes the subgroup of totally positive elements of K^* . The right hand side is a finite, étale scheme over $\mathbb{Z}_{(p)}$. The FROBENIUS element over K_p acts on it as the idèle $(1, \dots, p, \dots, 1)$ which is the image of p under the canonical immersion $K_p \subset K^*(\mathbb{A}_f)$. This coincides with the action of $(p^{-1}, \dots, p^{-1}) \in K^*(\mathbb{A}_f^p)$.

Let F_p be the unramified extension of degree 2 of K_p . We define $\tilde{\mathfrak{M}}_{B^*, \bar{C}}$ to be the form of $\mathfrak{M}_{B^*, \bar{C}}$ over $\mathbb{Z}_{(p)}$ on which the Frobenius element over \mathbb{F}_p acts as $\tau \circ (1, \dots, p^{-1}, \dots, 1)$, where τ denotes the action of the Frobenius element on $\mathfrak{M}_{B^*, \bar{C}}$. We get a morphism

$$(3) \quad \tilde{\mathfrak{M}}_{B^*, \bar{C}} \rightarrow Nm\bar{C} \backslash K^*(\mathbb{A}_f)/K_+^*.$$

The effect is, that we have now a constant scheme on the right hand side.

Let D be the quaternion algebra over K obtained by twisting B at p and the ramified infinite prime and D_- the quaternion algebra obtained by twisting B

at p and the unramified infinite prime. We fix isomorphisms

$$D(\mathbb{A}_f^p) = D_-(\mathbb{A}_f^p) = B(\mathbb{A}_f^p)$$

$$D_- \otimes_{\mathbb{Q}} \mathbb{R} = \mathbb{H} \times \mathbb{H}, \quad D \otimes_{\mathbb{Q}} \mathbb{R} = M_2(\mathbb{R}) \times M_2(\mathbb{R}).$$

Therefore we may consider C^p as a subgroup of $D^*(\mathbb{A}_f^p)$ and $D_-^*(\mathbb{A}_f^p)$ respectively. Let $C_p \subset D^*(\mathbb{Q}_p)$ and $C_{-p} \subset D_-^*(\mathbb{Q}_p)$ be the maximal compact subgroups. We define

$$C = C_p C^p \subset D^*(\mathbb{A}_f) \quad C_- = C_{-p} C^p \subset D_-^*(\mathbb{A}_f).$$

The reduced norm induces a map

$$C_- \backslash D_-^*(\mathbb{A}_f) / D_-^* \xrightarrow{Nm} NmC \backslash K^*(\mathbb{A}_f) / K_+^* = \pi_0(\mathfrak{M}_{B^*,C}).$$

Let $E \in \pi_0(\mathfrak{M}_{B^*,C})$ be a connected and hence a geometrically connected component. Let

$$k(E) = |Nm^{-1}(E)|$$

be the number of elements in the inverse image of E .

Theorem. *Let E be a connected component of $\mathfrak{M}_{B^*,C}$. If $p_a(E)$ denotes the genus of the curve E , we have*

$$p_a(E) = 1 + k(E)/(p^2 - 1), \quad p_a(E)/|E(\mathbb{F}_{p^4})| \leq 1/k(E) + 1/(p^2 - 1).$$

LANGLANDS [10] and ZINK [11] have considered the reduction of the SHIMURA variety $M_{D^*,C}$. There is a projective, flat morphism

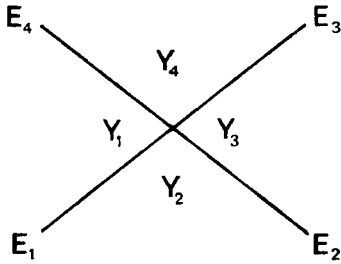
$$\mathfrak{M}_{D^*,C} \rightarrow \text{Spec } \mathbb{O}_{F_p}$$

such that the general fibre is $M_{D^*,C}$. Let $\tilde{\mathfrak{M}}_{D^*,C}$ be the scheme over \mathbb{O}_{F_p} obtained by twisting $\mathfrak{M}_{D^*,C}$ with the idèle $(1, \dots, p^{-2}, \dots, 1)$ as above. Then again $\pi_0(\tilde{\mathfrak{M}}_{D^*,C})$ is a constant scheme. The STEIN factorization is

$$\tilde{\mathfrak{M}}_{D^*,C} \rightarrow \pi_0(\tilde{\mathfrak{M}}_{D^*,C}) = NmC \backslash K^*(\mathbb{A}_f) / K_+^* = \pi_0(\mathfrak{M}_{B^*,C}).$$

Let E be a connected component of $\mathfrak{M}_{B^*,C}$ and X be the corresponding connected component of $\tilde{\mathfrak{M}}_{D^*,C}$. The special fibre Y of X is a union of 4 ruled surfaces

$$Y = \bigcup_{i \in \mathbb{Z}/4\mathbb{Z}} Y_i$$



The $E_i = Y_i \cap Y_{i+1}$ are smooth irreducible curves, which are all universally homeomorphic to the curve E . The intersections $Y_i \cap Y_{i+2} = Z$ coincide. Z is the constant scheme $Nm^{-1}(E)$. The formal equations of X at a point of Z are

$$x_1 y_1 = p, \quad x_2 y_2 = p.$$

Outside Z the divisor Y on X has normal crossings.

We denote by L_i one of the rational curves, which define the ruling on Y_i . On has following intersection numbers on Y_i .

$$(4) \quad (L_i \cdot E_i) = 1, \quad (L_i \cdot E_{i-1}) = p.$$

The Lemma 1 of KULIKOV [8] applied to our situation gives

$$(E_i^2)_{Y_{i+1}} + (E_i^2)_{Y_i} = 0.$$

By symmetry we have

$$(E_i^2)_{Y_{i+1}} = (E_{i-1}^2)_{Y_i}$$

and therefore

$$(5) \quad (E_{i-1}^2)_{Y_i} + (E_i^2)_{Y_i} = 0.$$

In the following computation all intersection products are taken on Y_i . We find by (4) the following relation in $\text{Num } Y_i$

$$E_{i-1} = pE_i + bL_i, \quad \text{for some } b \in \mathbb{Z}.$$

Multiplying this equation by E_i and E_{i-1} we get

$$p(E_i^2) + b = k, \quad pk + bp = (E_{i-1}^2),$$

where $k = k(E) = (E_i \cdot E_{i-1})$. Hence we obtain

$$(6) \quad p^2(E_i^2) + (E_{i-1}^2) = 2pk.$$

Let $K_i = xE_i + yL_i$, $x, y \in \mathbb{Z}$ be the canonical class on Y_i . Since $L_i = \mathbb{P}^1$, we find by the adjunction formula $p_a(C) = 1 + (C \cdot (C + K_i))/2$, that $x = -2$. By symmetry we have

$$\begin{aligned} p_a(E_i) &= p_a(E_{i-1}) \\ (E_i^2) + (E_i \cdot K_i) &= (E_{i-1}^2) + (E_{i-1} \cdot K_i) \\ (E_i^2) - 2(E_i^2) + y &= (E_{i-1}^2) - 2k + py. \end{aligned}$$

We deduce by (5)

$$y = 2k / (p - 1).$$

Using (5) and (6) we find $(E_i^2) = 2pk / (p^2 - 1)$ and

$$(7) \quad p_a(E_i) = 1 + (y - (E_i^2)) = 1 + k / (p^2 - 1).$$

This proves the first assertion of our theorem. The second follows because Z is a subscheme of E_i that consists of k points rational over \mathbb{F}_p .

§ 3. Let us fix a finite field \mathbb{F}_q and a smooth irreducible algebraic curve X of genus g over it. We suppose that X has n points Q, P_1, \dots, P_{n-1} defined over \mathbb{F}_q . We consider the linear space $\Omega(\sum P_i - \alpha Q)$ of differentials on X vanishing with multiplicity α at Q , regular everywhere outside the points P_i and having poles at most of the first order at the P_i . This space is isomorphic to $H^0(X, \mathcal{O}_X(K + \sum P_i - \alpha Q))$, where K is a canonical divisor on X . Its dimension is by the theorem of RIEMANN-ROCH at least $g + n - \alpha - 2$. Taking the residues at the points P_i we get a map

$$\begin{aligned} \Omega(\sum P_i - \alpha Q) &\rightarrow \mathbb{F}_q^{n-1} . \\ \omega &\mapsto (\text{Res}_{P_1} \omega, \dots, \text{Res}_{P_{n-1}} \omega) . \end{aligned}$$

The kernel of this map is $\Omega(-\alpha Q)$. It vanishes for $\alpha > 2g - 2$. The image of this map we call a GOPPA code (cf. [1]).

The advantage of this construction is that the basis in \mathbb{F}_q^{n-1} is almost canonical. It corresponds to \mathbb{F}_q -points of X . The minimal distance may be estimated by the theorem of RIEMANN-ROCH. In fact, let $D_1 \equiv \sum P_i$ be such a divisor that $\Omega(D_1 - \alpha Q) \neq 0$, i.e. such that there exists a differential in $\Omega(\sum P_i - \alpha Q)$ having no poles outside D_1 . Then $H^0(X, \mathcal{O}_X(K + D_1 - \alpha Q)) \neq 0$ and hence

$$2g - 2 + \deg D_1 - \alpha = \deg(K + D_1 - \alpha Q) \geq 0 .$$

The minimal distance d is therefore at least $\alpha - 2g + 2$.

Let us now consider the asymptotic of these codes when $n, g, \alpha \rightarrow \infty$. We have

$$R \equiv \frac{g + n - \alpha - 2}{n - 1} \approx \gamma + 1 - \frac{\alpha}{n}$$

$$\delta \equiv \frac{\alpha - 2g + 2}{n - 1} \approx \frac{\alpha}{n} - 2\gamma ,$$

where $\gamma = \frac{g}{n}$.

Choosing suitable $\alpha > 2g - 2$ we get codes with arbitrary δ and R , such that $\delta + R$ is asymptotically greater than $1 - \gamma$. Thus from every series of curves constructed in the previous paragraphs we get codes with $\gamma < 1$.

Let us now compare these codes with VARSHAMOV-GILBERT bound. First we compute when the GOPPA line $R = 1 - \gamma - \delta$ is tangent to the VARSHAMOV-GILBERT curve $R = 1 - \varphi(\delta)$.

$$\varphi(x) = x \log_q (q - 1) - x \log_q x - (1 - x) \log_q (1 - x)$$

$$\varphi'(x) = \log_q (q - 1) - \log_q x + \log_q (1 - x) .$$

We obtain $\varphi'(\delta_0) = 1$, when $\delta_0 = (q - 1) / (2q - 1)$. Consequently the GOPPA line is tangent if

$$\gamma = \varphi(\delta_0) - \delta_0 = \log_q (2q - 1) - 1 .$$

In other words, whenever $\gamma < \log_q(2q-1)-1$, there is an interval (δ_1, δ_2) inside of which GOPPA codes are better than those of VARSHAMOV-GILBERT. In the previous paragraphs we have found curves for $q=p^2$, $p \geq 7$ and for $q=p^4$, $p \geq 3$ with $\gamma \approx (\sqrt{q}-1)^{-1}$. Let $\delta_1 < \delta_2$ be the roots of $\varphi(\delta) - \delta = (\sqrt{q}-1)^{-1}$. We get the following.

Theorem. *GOPPA codes arising from reductions of the curves $X_0(l)$ over \mathbb{F}_q , $q=p^2$, $p \geq 7$ and from certain forms of reductions of SHIMURA curves over \mathbb{F}_q , $q=p^4$, $p \geq 3$ are better than VARSHAMOV-GILBERT codes in the intervall (δ_1, δ_2) .*

Added in proof. Yu. I. MANIN informed us that IHARA recently proved the following asymptotic inequality for γ if $q=p^{2m}$: $(\sqrt{q}-1)^{-1} \geq \gamma > (\sqrt{2q})^{-1}$.

References

- [1] ГОППА В. Д., Коды на алгебраических кривых. Проблемы передачи информации. 1982.
- [2] DELIGNE, P., Rapoport M. Les schémas de modules de courbes elliptiques. Lect. Notes in Math. N 849 (1973) 143–316.
- [3] DELIGNE, P., Travaux de Shimura Sémin. Bourbaki, année 1970/71, n° 389.
- [4] IHARA, Y., On congruence monodromy problems, v. 2. Tokyo 1968.
- [5] IHARA Y. On modular curves over finite fields. in "Discrete subgroups of Lie groups and applications to moduli", Oxford 1975.
- [6] IHARA Y., Congruence relations and Shimura curves. in Proc. Symp. Pure Math. 33, part 2, (1979) 291–311.
- [7] КАСАМИ Т., Токура Н., Ивадари Ё., Инагаки Я. Теория кодирования. Мир, Москва 1978.
- [8] Куликов В. С. Вырождение КЗ поверхностей Энриквеса. Изв. АН СССР, Сер. мат., 41 № 5, (1977) 1008–1042.
- [9] LANG S., Elliptic Functions. Reading. 1973.
- [10] LANGLANDS R. P., Sur la mauvaise reduction d'une variété de Shimura, Journée de Géom. Alg. de Rennes 1978, astérisque 65, 125–154.
- [11] ZINK, T. Über die schlechte Reduktion einiger Shimuramannigfaltigkeiten, Comp. Math. 1981.

Moscow State University

Akademie der Wissenschaften der DDR
Inst. für Mathematik
DDR-1086 Berlin
Mohrenstraße 39