MATHEMATICS OF COMPUTATION
VOLUME 38, NUMBER 157
JANUARY 1982

On Effective Computation of Fundamental Units. I

By Michael Pohst* and Hans Zassenhaus

Abstract. The new method for efficient computation of the fundamental units of an algebraic number field developed by the authors in an earlier paper is considerably improved with respect to (Section 1) utilization to best advantage of the element of choice inherent in the method and the mastery of the linear programming techniques involved, (Section 2) ideal factorization, and (Section 3) the determination of sharper upper bounds for the index of U_e in U_F.

Introduction. In [3] the authors devised a new method for efficient computation of the fundamental units of an algebraic number field. However, a few questions regarding the ensuing algorithm were left unanswered. We therefore announced further investigations. In the last four years we obtained much more computational experience with our method, and we discovered that some of the procedures could be considerably improved. The results of the theoretical analysis are contained in this paper, numerical examples and a description of the computer program will follow subsequently.

In Sections 1–3 we discuss the essential points of the computation of fundamental units in number fields, in accordance with the usual procedure. Sections 1 and 2 contain the determination of a maximal system of independent units, and in Section 3 we obtain an upper bound for the index of the group generated by the computed units in the full unit group. Then the latter can be computed without further theoretical difficulties as outlined in [3]. We use the following notation. F is an algebraic number field of degree $n \geqslant 2$ over $\mathbf{Q}$ with discriminant d_F. The ring of integers of F is denoted by o_F, its unit group by U_F. We assume that F has r_1 real and $2r_2$ complex conjugates, so that U_F contains $r = r_1 + r_2 - 1$ independent units according to the Dirichlet theorem.

In Section 1 we study the generation of arbitrarily many integers of F of bounded norm. We obtain several important improvements over [3] concerning the choice of a basis of o_F, the choice of ω for transforming the fundamental parallelotope and especially the determination of the lattice points in the transformed parallelotope.

In Section 2 we give a new method for the construction of units based on a sufficiently large number of algebraic integers of bounded norm. For real quadratic fields it was already discussed in [4].

By application of Sections 1 and 2 often enough we will obtain r independent units $\varepsilon_1, \ldots, \varepsilon_r$ and thus a system of generators of a subgroup U_e of finite index in

Received April 14, 1981.
1980 *Mathematics Subject Classification.* Primary 12A45.
* Author supported by the Deutsche Forschungsgemeinschaft.

0025-5718/82/0000-0490/$05.25

U_F. We do not deal here with the computation of a generating element of the torsion subgroup TU_F of U_F (for this see [3]). In order to determine a system of fundamental units derived from $\varepsilon_1, \ldots, \varepsilon_r$, we need an estimate of $(U_F: U_\varepsilon)$ and a method of extracting roots from suitable power products $\varepsilon_1^{m_1} \cdots \varepsilon_r^{m_r}$ $(m_i \in \mathbf{Z})$.

Improved upper bounds for the index of U_ε in U_F will be given in Section 3. Besides making use of conventional lower bounds for the regulator R_F of F we develop new lower estimates derived from Minkowski's Theorem on Successive Minima. Though this method was already outlined in [3], we present it anew with substantially stronger results.

The computation of U_F from U_ε—provided that an upper bound for $(U_F: U_\varepsilon)$ is known—was completely described in [3].

Part II contains a description of the corresponding computer program as well as an extensive list of numerical results obtained so far. They cover all fields of degrees $\leqslant 6$ and small absolute discriminant.

1. Construction of Integers of Small Norms. For any integral basis $\omega_1, \ldots, \omega_n$ of the given algebraic number field F there is a 1-1 correspondence between o_F and $\mathbf{Z}^n$ by means of the mapping

$$\varphi: F \to \mathbf{Q}^n: x_1\omega_1 + \cdots + x_n\omega_n \to (x_1, \ldots, x_n). \tag{1.1}$$

Therefore we can compute integers of F of bounded norms as lattice points in the parallelotope

$$\pi := \{(x_1, \ldots, x_n) \in \mathbf{Z}^n \mid -1 \leqslant x_i \leqslant 1, i = 1, \ldots, n\}, \tag{1.2}$$

respectively, in suitable linear transforms of π of equal volume. This procedure was described in detail in [3].

However, we did not discuss the effect which a particular choice of the basis $\omega_1, \ldots, \omega_n$ has on our method. To compute units from the integers of F obtained by the algorithm we would like their norms to be as small as possible. An upper bound is easily derived from (1.2). For

$$UB := \max\left\{\left|\prod_{j=1}^{n}\left(\sum_{i=1}^{n} x_i\omega_i^{(j)}\right)\right| \;\middle|\; -1 \leqslant x_i \leqslant 1, i = 1, \ldots, n\right\}, \tag{1.3}$$

where the product is taken over all conjugates, obviously

$$|N(\xi)| \leqslant UB \tag{1.4}$$

holds for every integer ξ of F which we obtain by our method. Therefore the integral basis $\omega_1, \ldots, \omega_n$ should be chosen so as to make UB as small as possible. Unfortunately the computation of such a basis is at least as difficult as the computation of units. For example, in real quadratic fields the corresponding extremal value problem would involve a solution of Pell's equation. Nevertheless, it is advisable to carefully consider the choice of $\omega_1, \ldots, \omega_n$ to make UB desirably small. A first step is, of course, the choice of basis elements of small norm. Since our method provides such elements at each step, a change of basis after a few steps is highly recommendable.

Example: $n = 2$, $F = \mathbf{Q}(\sqrt{6})$, $o_F = \mathbf{Z} + \sqrt{6}\mathbf{Z}$. For $1, \sqrt{6}$ as basis elements we obtain $UB = 6$. If we choose $1, 2 + \sqrt{6}$ instead, we get $UB = 5$. Moreover, π then contains a lattice point corresponding to $3 + \sqrt{6}$. After we change the basis to

$\omega_1 = 2 + \sqrt{6}$, $\omega_2 = 3 + \sqrt{6}$, we obtain firstly $UB = 3$, and then also the fundamental unit $\varepsilon = \omega_1 + \omega_2$ of F as a lattice point of π. Hence we can compute the fundamental unit in that case without transforming π at all.

In most cases, however, ω-transformations of π cannot be avoided. For $\omega \in o_F \setminus \mathbf{Z}$ we compute the right regular representation matrix M_ω by means of

$$\omega(\omega_1, \ldots, \omega_n) = (\omega_1, \ldots, \omega_n)M_\omega. \tag{1.5}$$

Then the matrix $|N(\omega)|^{-1/n}M_\omega$ yields a linear transformation of determinant ± 1. Denote the transformed parallelotope by $\pi(\omega)$:

$$\pi(\omega) = \left\{|N(\omega)|^{-1/n}(x_1, \ldots, x_n)M_\omega^t \mid -1 \leqslant x_i \leqslant 1, i = 1, \ldots, n\right\}. \tag{1.6}$$

From this presentation it is easily seen that the bound UB of (1.3) remains unaltered for $\pi(\omega)$.

An optimal choice for ω is still an unsolved problem. We propose to call ω "good", if $\pi(\omega)$ contains many lattice points different from those of π. Unfortunately we do not know how to make the best choice of ω to that effect. So far we can only give rules of thumb derived from our experience with extensive calculations. New lattice points rarely occurred if we chose ω at random or if we made one coefficient in the representation of ω by $\omega_1, \ldots, \omega_n$ large and the other coefficients small. Substantially better results were obtained by using a number ω of small norm and continuing by using a few consecutive powers of that same number. Apart from those improved results we gained also the following advantages. Firstly, numbers of small norms are already derived from the lattice points in π, they are always at hand. So we do not need an extra routine for the generation of ω, we can simply choose the element of smallest norm in storage which was not yet used for this purpose. Secondly, we must compute M_ω only once, compare the remarks to Step (A) below. Further simplifications can occur in the process of determining lattice points of $\pi(\omega)$. We shortly recall the main steps from [3].

(A) By the elementary divisor theorem a unimodular matrix U and its inverse are computed such that $MU =: N$ is a lower triangular matrix.

(B) A lower triangular matrix $B \in \mathbf{Z}^{n \times n}$ is recursively computed from $NB = |N(\omega)|I_n$, I_n denoting the $n \times n$ unit matrix.

(C) The integral solutions $\mathbf{x} \in \mathbf{Z}^n$ of the system of inequalities

$$-|N(\omega)|^{(n-1)/n}(1, \ldots, 1)^t \leqslant \mathbf{x}B \leqslant |N(\omega)|^{(n-1)/n}(1, \ldots, 1)^t \tag{1.7}$$

can easily be calculated, since in the ith inequality only $x_n, \ldots, x_i$ occur ($i = n, n-1, \ldots, 1$).

(D) Every lattice point $\mathbf{y}$ of $\pi(\omega)$ is obtained from a solution $\mathbf{x}$ of (C) by means of $\mathbf{y} = \mathbf{x}U^{-1}$.

In the sequel we investigate steps (A) and (C) for further improvements.

Step (A). When we transform π by successive powers of an integer ω of F, we can avoid computing $M_{\omega^\nu} = (M_\omega)^\nu$ ($\nu = 2, 3, \ldots$) at each step. For the determination of the lattice points of the transformed parallelotope we do not actually need M_{ω^ν} but only the lower triangular matrix N_{ω^ν} obtained from it by column operations and multiplication by the corresponding unimodular matrices U_{ω^ν} and $U_{\omega^\nu}^{-1}$, respectively. We therefore compute those for consecutive powers of ω in the following

way. Because of $M_\omega = N_\omega U_\omega^{-1}$ and $M_\omega^\nu = M_\omega M_{\omega^{\nu-1}} = (M_\omega N_{\omega^{\nu-1}}) U_{\omega^{\nu-1}}^{-1}$ for $\nu = 2, 3, \ldots$, we calculate for each $\nu \geqslant 2$ the matrix $M_\omega N_{\omega^{\nu-1}}$ which cuts down the work by nearly one half, since $N_{\omega^{\nu-1}}$ is already a lower triangular matrix. Thereafter $M_\omega N_{\omega^{\nu-1}}$ is transformed into a lower triangular matrix N_{ω^ν} and at each step $U_{\omega^{\nu-1}}^{-1}$ is transformed analogously yielding $M_\omega^\nu = N_{\omega^\nu} U_{\omega^\nu}^{-1}$. Moreover, we can give N_ω (N_{ω^ν} respectively) a canonically reduced form by column operations demanding for its entries n_{ij}:

$$(1.8) \qquad n_{ii} > 0, \quad -\tfrac{1}{2} n_{ii} < n_{ij} \leqslant \tfrac{1}{2} n_{ii} \qquad (i = 2, \ldots, n; j = 1, \ldots, i-1).$$

Step (C). Solving the system of linear inequalities in the proposed way is recommendable only in case $|N(\omega)|^{(n-1)/n}$ is small compared with the entry b_{nn} of B. Usually b_{nn} will be one. So, if $|N(\omega)|$ becomes large, for example, when ω is a power of some integer of F, we would have to consider the remaining $n-1$ inequalities for a huge number of values x_n, most of which would not yield a solution of the system. We therefore proceed in either of two ways.

Method 1. We compute $w := \max\{b_{jj}|N(\omega)|^{(1-n)/n} \mid j = 1, \ldots, n\}$ and choose i maximal with the property $b_{ii}|N(\omega)|^{(1-n)/n} = w$. Then we eliminate $x_n, x_{n-1}, \ldots, x_{i+1}$ from the ith inequality by adding suitable multiples of the inequalities $n, n-1, \ldots, i+1$ to it. This yields a much smaller number of possibilities for x_i and consecutively we determine all solutions $(x_i, \ldots, x_n)$ for the inequalities $i, i+1, \ldots, n$. In order not to obtain superfluous solutions we must store all inequalities which occur during the elimination. Fortunately the size of the problems we considered so far allowed this without difficulties. For the $(n-i+1)$-tuples $(x_i, \ldots, x_n)$ obtained in that way we then compute $x_{i-1}, \ldots, x_1$ (if existing) as suggested in (C).

Method 2. This method should be applied in case the matrix B computed from canonically reduced N's is sparse below the diagonal. For example, if $|N(\omega)|$ is a prime number, B only contains entries different from zero in the last row and—of course—in the diagonal. Hence, only the variables x_i, x_n occur in the ith inequality of (1.7), and instead of (1.7) we obtain

$$(1.9) \qquad \begin{aligned} -|N(\omega)|^{(n-1)/n} \leqslant b_{ii}x_i + b_{in}x_n \leqslant |N(\omega)|^{(n-1)/n}, \\ 1 \leqslant x_n \leqslant |N(\omega)|^{(n-1)/n}/b_{nn} \end{aligned}$$

$(i = 1, \ldots, n-1)$. We substitute x_n by $x_n - 1$, divide the first inequality by $\gcd(b_{ii}, b_{in})$, and replace x_n by $|N(\omega)|^{(n-1)/n}/b_{nn} - 1 - x_n$ in case $b_{in} < 0$. Then it suffices to determine the lattice points of a parallelogram given by

$$(1.10) \qquad j \leqslant ax + by \leqslant k, \qquad 0 \leqslant y \leqslant l \; (j \in \mathbf{Z}, k, l, a, b \in \mathbf{N}; x, y \in \mathbf{Z}^{>0}).$$

This can be done very efficiently by a linear program:

$$(1.11) \qquad \max(ax + by) \text{ subject to } ax + by \leqslant k, \qquad 0 \leqslant y \leqslant l, x \geqslant 0.$$

The solution of (1.11) will be a lattice point (x_0, y_0) satisfying $j \leqslant ax_0 + by_0 \leqslant k$, because we know from Minkowski's Convex Body Theorem that there is at least one lattice point different from (0, 0). In order to obtain further lattice points satisfying (1.10) we replace k by $Ax_0 + By_0 - 1$ in (1.11). This procedure can be carried out as long as the optimal solution of (1.11) yields a value which is greater or equal to j.

In general it takes only a few steps to solve (1.11). From $ax + by \leqslant k$ we derive $y \leqslant \min\{[k/b], l\}$. In case $k/b \leqslant l$, we can drop the condition $y \leqslant l$ and go to (1.13) setting $m = 0$. In the other case: $k/b > l$ we decompose the pointset $\{(x, y) \in (\mathbf{R}^{\geqslant 0})^2 \mid ax + by \leqslant k, y \leqslant l\}$ into a rectangle $\{(x, y) \in (\mathbf{R}^{\geqslant 0})^2 \mid 0 \leqslant y \leqslant l,\ 0 \leqslant x \leqslant (k - bl)/a\}$ and a rectangular triangle with vertices $((k - bl)/a, 0)$, $((k - bl)/a, l)$, $(k/a, 0)$. The optimal value of $ax + by$ on the rectangle is obtained for $x = [(k - bl)/a]$, $y = l$. It remains to discuss the lattice points of the triangle. We set

$$m := \begin{cases} (k - bl)/a & \text{if it is in } \mathbf{Z}, \\ [(k - bl)/a] + 1 & \text{else,} \end{cases} \tag{1.12}$$

and substitute $x - m$ for x. This yields

$$\max(ax + by + am) \text{ subject to } ax + by \leqslant k - am =: \bar{k}, \quad x \geqslant 0, y \geqslant 0. \tag{1.13}$$

Without loss of generality we can assume $a > b > 0$. For an optimal solution of (1.13) we must necessarily have $y \geqslant [(\bar{k} - ax)/b]$. Hence we can lessen the pointset under consideration by applying Euclid's algorithm [1]. Let $q, r \in \mathbf{Z}$ such that $a = qb + r$ and $0 \leqslant r < b$. Then

$$y \geqslant \left[\left(\frac{\bar{k}}{a} - x\right)\frac{a}{b}\right] \geqslant \left[\frac{\bar{k}}{a} - x\right]\left[\frac{a}{b}\right] = \left(\left[\frac{\bar{k}}{a}\right] - x\right)q,$$

and $x \leqslant [\bar{k}/a]$. For $[\bar{k}/a] = 0$ we are done with $x = 0$, $y = [k/b]$ as an optimal solution. Else we substitute $y - ([\bar{k}/a] - x)q$ for y and obtain

$$\max\left((a - bq)x + by + am + \left[\frac{\bar{k}}{a}\right]qb\right) \text{ subject to}$$

$$(a - bq)x + by \leqslant \bar{k} - \left[\frac{\bar{k}}{a}\right]qb,\ 0 \leqslant x \leqslant \left[\frac{\bar{k}}{a}\right], y \geqslant 0. \tag{1.14}$$

It is easily seen that (1.13) and (1.14) are equivalent in the sense that their optimal solutions coincide. On the other hand (1.14) is a problem of the same form as (1.11) but with much smaller values instead of a, k, l. After a few transformations of this kind the process comes to a halt because of either one of the numbers a, b, k, l being zero or $a > k$ or $b > k$. For an estimate of the number of necessary steps see [1]. We only mention that the solution is obtained in a polynomial number of steps depending on the size of a, b.

Method 2 can be strongly recommended if we can either assume that there are not too many lattice points, or if we only want to determine a few of them.

2. On the Factorization of Nonzero Principal o_F-Ideals.

Introduction. Given a finite number of nonzero elements $\alpha_1, \alpha_2, \ldots, \alpha_m$ of o_F, we want to factorize the principal ideals $\alpha_i o_F$ into power products

$$\alpha_i o_F = \prod_{j=1}^{t} \prod_{k=1}^{\tau_j} \mathfrak{a}_{jk}^{\nu_{jk}(\alpha_i)} \qquad (1 \leqslant i \leqslant m) \tag{2.1}$$

of mutually prime ideals of the form

$$\mathfrak{a}_{jk} := a_j o_F + \alpha_{jk} o_F \qquad (1 \leqslant k \leqslant \tau_j,\ 1 \leqslant j \leqslant t) \tag{2.2a}$$

of o_F that are distinct from o_F such that

(2.2b) $$t, \tau_1, \tau_2, \ldots, \tau_t, a_1, \ldots, a_t$$

are natural numbers and

(2.2c) $$\gcd(a_i, a_k) = 1 \qquad (1 \leqslant i < k \leqslant t),$$

the elements α_{jk} are contained in o_F and the exponents $\nu(\alpha_i, \alpha_{jk})$ are nonnegative rational integers.

I. *Contents.* In order to solve the task stated in the introduction, we introduce the content concept. For the purpose of computing contents, it is convenient to avail oneself of a minimal basis of F, i.e., of a $\mathbf{Z}$-basis $\omega_1, \ldots, \omega_n$ of o_F.

Definition (2.3). The (rational) *content* of an element

(2.4) $$\gamma = \sum_{i=1}^{n} e_i\omega_i \qquad (e_i \in \mathbf{Q},\ 1 \leqslant i \leqslant n)$$

of F is defined as the greatest common divisor of the coefficients $e_1, \ldots, e_n$:

(2.5) $$C(\gamma) := \gcd(e_1, \ldots, e_n), \qquad 0 \leqslant C(\gamma) \in \mathbf{Q}, \qquad C(\gamma)\mathbf{Z} = \sum_{i=1}^{n} e_i\mathbf{Z}.$$

For example the content of a rational number e is $|e|$.

LEMMA (2.6). *For any element γ of F the fractional o_F-ideal $C(\gamma)o_F$ generated by the content of γ is the intersection of the fractional $\mathbf{Z}$-ideals $\lambda\mathbf{Z}$ satisfying the condition*

(2.7) $$\lambda o_F \ni \gamma.$$

Proof. (2.4), (2.5) imply that

$$e_i = C(\gamma)x_i \qquad (x_i \in \mathbf{Z},\ 1 \leqslant i \leqslant n),$$

and therefore

$$\gamma = C(\gamma)\sum_{i=1}^{n} x_i\omega_i \in C(\gamma)o_F = (C(\gamma)\mathbf{Z})o_F.$$

Conversely, let $\lambda\mathbf{Z}$ be a fractional $\mathbf{Z}$-ideal satisfying (2.7). Then we have $e_i \in \lambda\mathbf{Z}$ and $C(\gamma)\mathbf{Z} = \sum_{i=1}^{n} e_i\mathbf{Z} \subseteq \lambda\mathbf{Z}$.

LEMMA (2.8). *For any nonzero element γ of F the content of γ is derived from the coefficients of the minimal polynomial $P(t) = P_\gamma(t)$:*

(2.9) $$P(t) = t^{n_\gamma} + \sum_{i=1}^{n_\gamma} a_i t^{n_\gamma - i},$$
$$P(\gamma) = 0 \qquad \left(n_\gamma \in \mathbf{Z}^{>0};\ a_i \in \mathbf{Q},\ 1 \leqslant i \leqslant n_\gamma;\ a_{n_\gamma} \neq 0\right)$$

as the largest rational number x satisfying

(2.10) $$a_i \in x^i\mathbf{Z} \qquad (1 \leqslant i \leqslant n_\gamma).$$

Proof. According to (2.4), (2.5), we have $\gamma = C(\gamma)\omega$ with ω in o_F. Hence the minimal polynomial of ω is

$$P_\omega(t) = t^{n_\gamma} + \sum_{i=1}^{n_\gamma} a_i C(\gamma)^{-i} t^{n_\gamma - i}.$$

Moreover, the coefficients of $P_\omega(t)$ are rational integers. Hence

$$a_i \in C(\gamma)^i\mathbf{Z} \qquad (1 \leqslant i \leqslant n_\gamma).$$

Conversely, let x be a rational number satisfying (2.10). It follows that $x \neq 0$ and that $x^{-1}\gamma$ belongs to o_F, hence $\gamma \in xo_F$. Now Lemma (2.6) yields that x belongs to $C(\gamma)\mathbf{Z}$, hence Lemma (2.8).

LEMMA (2.11). *For any element γ of F and for any rational number x, we have*

$$C(x\gamma) = |x|C(\gamma). \tag{2.12}$$

LEMMA (2.13). *For any two elements γ_1, γ_2 of F, we have*

$$C(\gamma_1 + \gamma_2) \in C(\gamma_1)\mathbf{Z} + C(\gamma_2)\mathbf{Z}, \tag{2.14a}$$

$$C(\gamma_1\gamma_2) \in C(\gamma_1)C(\gamma_2)\mathbf{Z}. \tag{2.14b}$$

The proofs of Lemmas (2.11), (2.13) are trivial.

As a consequence of (2.14) and the definition (2.6), we have

$$C(\beta^{-1})^{-1}\mathbf{Z} = \mathbf{Q} \cap \beta o_F \qquad (0 \neq \beta \in F). \tag{2.15}$$

Finally we generalize our content concept to

Definition (2.16). Let F be a commutative unital ring with subring R satisfying the conditions

(a) $1_F \in R$,

(b) any nonzero divisor of R is a nonzero divisor of F.

Furthermore let S be an overring of R in F such that F is the quotient ring of S.

It follows that F contains the quotient ring $Q = Q(R)$ of R. The *S/R-content ideal* of a fractional S-ideal γ contained in F is defined as the intersection $C_{S/R}(\gamma)$ of all fractional R-ideals $\mathfrak{a}$ of Q satisfying $\gamma \subseteq \mathfrak{a}S$.

THEOREM (2.17). *In the situation of Definition* (2.16) *we have*

$$C_{S/R}(xS) = xR \qquad (x \in Q); \tag{2.18}$$

$$C_{S/R}(x\gamma) = xC_{S/R}(\gamma) \qquad (x \text{ any unit of } Q); \tag{2.19}$$

$$C_{S/R}(\gamma_1 + \gamma_2) \subseteq C_{S/R}(\gamma_1) + C_{S/R}(\gamma_2); \tag{2.20}$$

$$C_{S/R}(\gamma_1\gamma_2) \subseteq C_{S/R}(\gamma_1)C_{S/R}(\gamma_2) \qquad (\gamma_1, \gamma_2 \text{ any two fractional } S\text{-ideals of } F); \tag{2.21}$$

(2.22) *if F is an integral domain and if R is integrally closed relative to Q, then for any algebraic element γ of S with minimal polynomial* (2.9) *with a_i in Q (instead of $\mathbf{Q}$) for $i = 1, 2, \ldots, n_\gamma$, the S/R-content of γS is the intersection of the fractional R-ideals $\mathfrak{a}$ of Q satisfying $a_i \in \mathfrak{a}^i$ $(1 \leqslant i \leqslant n_\gamma)$;*

(2.23) *if γ is an invertible fractional S-ideal of F and R is a Dedekind domain, then $C_{S/R}(\gamma^{-1})^{-1} = \gamma \cap Q$.*

Proof. Generalize the proofs of Lemmas (2.6), (2.8), (2.11), (2.13), and the proof of (2.15).

Returning to the initial situation we describe the connection between the content of an element γ of o_F and the $F/\mathbf{Q}$ norm of γ as follows.

Since $C(\gamma)$ is a rational integer dividing γ in o_F it follows that $C(\gamma)^n \mid N(\gamma)$, where $N(\gamma) = N_{F/\mathbf{Q}}(\gamma)$ is the principal norm of γ over $\mathbf{Q}$. On the other hand let

$$\gamma^n + a_1\gamma^{n-1} + \cdots + (-1)^n N(\gamma) = 0$$

be the principal equation of γ over $\mathbf{Q}$, so that $-a_1 = \mathrm{tr}(\gamma) = \mathrm{tr}_{F/\mathbf{Q}}(\gamma)$ is the principal trace of γ over $\mathbf{Q}$ and $a_1, a_2, \ldots, a_n = (-1)^n N(\gamma)$ are rational integers. If $N(\gamma) = 0$, then $\gamma = C(\gamma) = N(\gamma) = 0$. If $N(\gamma) \neq 0$, then we use the equation

$$\tilde{\gamma}\gamma = N(\gamma), \qquad \tilde{\gamma} = (-1)^{n+1}\left(\gamma^{n-1} + a_1\gamma^{n-2} + \cdots + a_{n-1}\right),$$

which implies the divisibility

$$C(\gamma^{-1})^{-1} \mid N(\gamma)\left(\gcd\left(C(\gamma)^{n-1}, C(\gamma)^{n-2}a_1, \ldots, a_{n-1}\right)^{-1}\right),$$

so that both $C(\gamma)^n$ and $C(\gamma^{-1})^{-1}$ are divisors of $N(\gamma)$.

II. *Normal Ideal Presentations.*

Definition (2.24). Given a fractional o_F-ideal $\mathfrak{a}$. Any presentation of $\mathfrak{a}$ of the form

$$\mathfrak{a} = \delta(ao_F + \alpha o_F) \tag{2.25a}$$

with

$$a \in \mathbf{Z}^{>0}, \quad \alpha \in o_F, \qquad 0 \neq \delta \in \mathbf{Q}, \tag{2.25b}$$

such that either

$$\alpha \neq 0, \quad C(\alpha^{-1})a = \zeta\varepsilon^{-1}, \qquad \zeta, \varepsilon \in \mathbf{Z}, \quad C_{o_F/\mathbf{Z}}(\mathfrak{a}^{-1})^{-1} = a\delta\mathbf{Z}, \quad \gcd(\zeta, \varepsilon) = \gcd(a, \varepsilon) = 1 \tag{2.25c}$$

or

$$\alpha = 0, \qquad \mathfrak{a} = \delta a o_F \tag{2.25d}$$

is said to be a *normal presentation* of $\mathfrak{a}$.

LEMMA (2.26). *For any fractional o_F-ideal $\mathfrak{a}$ there is a normal presentation.*

Proof. If $\mathfrak{a} = 0$, then we obtain a fractional presentation (2.25a) upon setting $\delta = 1$, $a = 0$, $\alpha = 0$.

Now let $\mathfrak{a} \neq 0$. Then

$$C_{o_F/\mathbf{Z}}(\mathfrak{a}^{-1})^{-1} = a\delta\mathbf{Z} \tag{2.27a}$$

with a, δ satisfying (2.25b) and

$$\delta^{-1} \in \mathbf{Z}^{>0}, \qquad \gcd(a, \delta^{-1}) = 1. \tag{2.27b}$$

Furthermore, we have $a\delta \in \mathfrak{a}$, therefore $a \in \delta^{-1}\mathfrak{a} \subseteq o_F$. But by a well-known lemma of ideal theory there is an element α of $\mathfrak{a}$ for which $\delta^{-1}\mathfrak{a} = ao_F + \alpha o_F$; hence (2.25a) is satisfied. We gather from (2.27a) that $C_{o_F/\mathbf{Z}}(\delta\mathfrak{a}^{-1})^{-1} = a\mathbf{Z}$; hence (2.25c) is satisfied in case $\alpha \neq 0$. Otherwise we have (2.25d).

Let us observe that there is the counterpart of (2.27a)

$$C_{o_F/\mathbf{Z}}(\mathfrak{a}) = \delta \gcd(a, C(\alpha))\mathbf{Z} \tag{2.27c}$$

valid for *every* normal presentation (2.25a) of $\mathfrak{a}$.

Example. For any nonzero principal ideal αo_F of o_F there holds the normal presentation (2.25a) with $\delta = 1$, $a = C(\alpha^{-1})^{-1}$ and (2.25c). This normal presentation of αo_F is said to be the *α-normal presentation.*

Definition (2.28). The normal presentation (2.25a) is *straightened out* upon transition to the normal presentation:

$$\mathfrak{a} = \delta'(a' o_F + \alpha' o_F), \tag{2.29a}$$

where

$$a'\mathbf{Z} = aC_{o_F/\mathbf{Z}}(ao_F + \alpha o_F)^{-1}, \tag{2.29b}$$

$$\alpha'\mathbf{Z} = \alpha C_{o_F/\mathbf{Z}}(ao_F + \alpha o_F)^{-1}, \tag{2.29c}$$

$$\delta'\mathbf{Z} = \delta C_{o_F/\mathbf{Z}}(ao_F + \alpha o_F), \tag{2.29d}$$

in case (2.25c). But in case $a \neq 0$, $\alpha = 0$ we straighten out (2.25a) to

$$\mathfrak{a} = \delta' o_F, \tag{2.29e}$$

where (2.29d) obtains again.

We observe that in either case

$$a' \mid a, \quad \alpha' \mid \alpha, \quad C_{o_F/\mathbf{Z}}(a' o_F + \alpha' o_F) = 1, \tag{2.29f}$$

$$C_{o_F/\mathbf{Z}}(\mathfrak{a}') = \delta'\mathbf{Z}, \tag{2.29g}$$

and we note that

$$C_{o_F/\mathbf{Z}}(ao_F + \alpha o_F) = \gcd(a, C(\alpha))\mathbf{Z}. \tag{2.29h}$$

The computational advantage of straightening out a normal presentation lies in the reduction of the size of the numbers a, α. The proof of Lemma (2.26) yields a straightened out normal presentation of $\mathfrak{a}$.

We observe that $\mathfrak{a} = o_F$ precisely if $\delta' = 1$, $a' = 1$.

LEMMA (2.30). *A normal presentation of the inverse of the ideal $\mathfrak{a} \neq 0$ normally presented according to* (2.25a–d) *is given by*

$$\mathfrak{a}^{-1} = \delta^{-1} \gcd\big(a, C(\alpha^{-1})^{-1}\big)^{-1}\big(ao_F + C(\alpha^{-1})^{-1}\alpha^{-1} o_F\big), \tag{2.31a}$$

in case (2.25c), *but*

$$\mathfrak{a}^{-1} = \delta^{-1} a^{-2}(ao_F + 0o_F), \tag{2.31b}$$

in case (2.25d).

Proof. The verification is straightforward.

LEMMA (2.32). *If $\mathfrak{a}$, $\mathfrak{a}'$ are two fractional o_F-ideals, both of which are normally presented, say $\mathfrak{a}$ by* (2.25a) *and $\mathfrak{a}'$ by*

$$\mathfrak{a}' = \delta'(a' o_F + \alpha' o_F) \tag{2.33a}$$

with

$$a' \in \mathbf{Z}^{>0}, \quad \alpha' \in o_F, \qquad 0 \neq \delta' \in \mathbf{Q}, \tag{2.33b}$$

such that either

$$\alpha' \neq 0, \quad C(\alpha'^{-1})a' = \zeta'\varepsilon'^{-1}, \qquad \zeta', \varepsilon' \in \mathbf{Z}, \quad C_{o_F/\mathbf{Z}}(\mathfrak{a}'^{-1})^{-1} = a'\delta'\mathbf{Z}, \tag{2.33c}$$
$$\gcd(\zeta', \varepsilon') = \gcd(a', \varepsilon') = 1$$

or

(2.33d) $$\alpha' = 0, \quad \mathfrak{a}' = \delta' a' o_F,$$

and that, moreover,

(2.33e) *the natural numbers a, a' share every prime divisor in case $\alpha \neq 0$, $\alpha' \neq 0$,*

then we obtain the normal presentation

(2.34a) $$\mathfrak{a}\mathfrak{a}' = \delta\delta'(aa' o_F + \alpha\alpha' o_F)$$

of the product ideal if $\alpha \neq 0$, $\alpha' \neq 0$. But if $\alpha = 0$, then we have instead

(2.34b) $$\mathfrak{a}\mathfrak{a}' = \delta\delta'(aa' o_F + a\alpha' o_F),$$

and if $\alpha' = 0$, then we take

(2.34c) $$\mathfrak{a}\mathfrak{a}' = \delta\delta'(aa' o_F + \alpha a' o_F)$$

as normal presentation of the product ideal.

Proof. The cases (2.34b–c) are trivial. Now let us deal with the case that $\alpha \neq 0$, $\alpha' \neq 0$. Then we assume $\delta = \delta' = 1$ without loss of generality, so that we have to show that

(2.34d) $$\mathfrak{a}\mathfrak{a}' = aa' o_F + \alpha\alpha' o_F$$

and that this presentation of the product ideal is normal.

The equation (2.34d) is a consequence of (2.25b–c), (2.33b–c). The normality of the presentation (2.34d) follows from (2.33e).

If $\mathfrak{a}$, $\mathfrak{a}'$ are two nonzero ideals with straightened out normal presentation (2.25a), (2.33a), respectively, then the equality of $\mathfrak{a}$, $\mathfrak{a}'$ is tantamount to $\delta = \delta'$, $a = a'$ and $\mathfrak{a}^{-1}\mathfrak{a}' = o_F$.

Moreover, the inclusion

(2.35a) $$\mathfrak{a} \supseteq \mathfrak{a}'$$

implies the necessary conditions

(2.35b) $$C_{o_F/\mathbf{Z}}(\mathfrak{a}) \supseteq C_{o_F/\mathbf{Z}}(\mathfrak{a}'),$$

(2.35c) $$C_{o_F/\mathbf{Z}}(\mathfrak{a}^{-1}) \subseteq C_{o_F/\mathbf{Z}}((\mathfrak{a}')^{-1})$$

that are equivalent to the divisibilities

(2.35d) $$\delta \mid \delta',$$

(2.35e) $$\delta a \mid \delta' a',$$

respectively.

In case $\mathfrak{a} = 0$, they imply $\mathfrak{a}' = 0$ and thus (2.35a).

But if $\mathfrak{a} \neq 0$, then (2.35a) is equivalent to

(2.35f) $$o_F \supseteq \mathfrak{a}^{-1}\mathfrak{a}'.$$

Suppose the two necessary conditions (2.35d), (2.35e) are fulfilled already; then (2.35f) is equivalent to the divisibility

(2.35g) $$a \,\Big|\, \frac{\delta'}{\delta} C(\alpha'),$$

in case $\alpha = 0$.

But in case $\alpha \neq 0$ we have

$$\gcd\left(a, C(\alpha^{-1})^{-1}\right) = a, \tag{2.35h}$$

$$\mathfrak{a}^{-1} = \delta^{-1}\left(o_F + a^{-1}C(\alpha^{-1})^{-1}\alpha^{-1}o_F\right), \tag{2.35i}$$

and then (2.35f) is equivalent to

$$a \mid \frac{\delta'}{\delta} C\left(\alpha' C(\alpha^{-1})^{-1}\alpha^{-1}\right). \tag{2.35j}$$

Thus we obtain a test for each of the five relations

$$\mathfrak{a} = \mathfrak{a}', \quad \mathfrak{a} \supseteq \mathfrak{a}', \quad \mathfrak{a}' \supseteq \mathfrak{a}, \quad \mathfrak{a} \supset \mathfrak{a}', \quad \mathfrak{a}' \supset \mathfrak{a},$$

which is computationally feasible. Using this remark and Lemma 2.32 we are able to refine any nontrivial ideal factorization so long until any two factors either are equal or they are mutually prime.

Now we will provide two ways of factoring given ideals into a product of proper divisors. Firstly, given a normal presentation

$$\mathfrak{a} = ao_F + \alpha o_F \tag{2.36a}$$

and a factorization

$$a = a_1 a_2 \tag{2.36b}$$

of the natural number a into the product of two mutually prime natural numbers greater than 1:

$$a_1 \in \mathbf{Z}^{>1}, \quad a_2 \in \mathbf{Z}^{>1}, \quad \gcd(a_1, a_2) = 1, \tag{2.36c}$$

then we have the ideal factorization

$$\mathfrak{a} = \mathfrak{a}_1\mathfrak{a}_2, \quad \mathfrak{a}_1 = a_1 o_F + \alpha o_F, \quad \mathfrak{a}_2 = a_2 o_F + \alpha o_F \tag{2.36d}$$

of $\mathfrak{a}$ into the product of two mutually prime normally presented ideals $\mathfrak{a}_1$, $\mathfrak{a}_2$.

In case the normal presentation of $\mathfrak{a}$ is straightened out, it follows that also the normal presentations of $\mathfrak{a}_1$, $\mathfrak{a}_2$ are straightened out and that both $\mathfrak{a}_1$, $\mathfrak{a}_2$ are properly contained in o_F.

Though it is computationally difficult to factorize a given nonzero fractional o_F-ideal $\mathfrak{a}$ into a power product of prime ideals of o_F, for theoretical purposes that factorization is of inestimable value. The possibility of such a factorization is the defining earmark of a Dedekind domain (like o_F). The uniqueness of the factorization can be shown to be a consequence of its universal existence.

Let us analyze the concepts of the content of $\mathfrak{a} \neq 0$ and derived concepts in the light of the prime ideal factorization of $\mathfrak{a}$ which we present in the form

$$C_{o_F/\mathbf{Z}}(\mathfrak{a}^{-1})^{-1} = \prod_{i=1}^{\sigma} p_i^{\nu(\delta a, p_i)}\mathbf{Z}, \tag{2.37a}$$

$$p_i o_F = \prod_{k=1}^{g_i} \mathfrak{p}_{ik}^{\nu(p_i, \mathfrak{p}_{ik})}, \tag{2.37b}$$

$$\mathfrak{a} = \prod_{i=1}^{\sigma}\prod_{k=1}^{g_i} \mathfrak{p}_{ik}^{\nu(\mathfrak{a}, \mathfrak{p}_{ik})}, \tag{2.37c}$$

where $p_1, p_2, \ldots, p_\sigma$ are distinct prime numbers, σ is a nonnegative rational integer, the prime ideals $\mathfrak{p}_{ik}$ $(1 \leqslant k \leqslant g_i)$ are the distinct prime ideal divisors of p_i in o_F (g_i a natural number $\leqslant n$) and furthermore

$$\mathbf{Z} \ni \nu(\delta a, p_i) \neq 0 \qquad (1 \leqslant i \leqslant \sigma); \tag{2.37d}$$

$$\nu(p_i, \mathfrak{p}_{ik}) \in \mathbf{Z}^{>0} \qquad (1 \leqslant k \leqslant g_i), \tag{2.37e}$$

$$\begin{aligned} \nu(\delta a, p_i) &= \nu\left(C_{o_F/\mathbf{Z}}(\mathfrak{a}^{-1})^{-1}, p_i\right) \\ &= \min_{1 \leqslant k \leqslant g_i} \left[-\nu(\mathfrak{a}, \mathfrak{p}_{ik})/\nu(p_i, \mathfrak{p}_{ik})\right] \qquad (1 \leqslant i \leqslant \sigma), \end{aligned} \tag{2.37f}$$

$$\begin{aligned} \nu\left(C_{o_F/\mathbf{Z}}(\mathfrak{a}), p_i\right) &= \min_{1 \leqslant k \leqslant g_i} \left[\nu(\mathfrak{a}, \mathfrak{p}_{ik})/\nu(p_i, \mathfrak{p}_{ik})\right] \\ &= \min(\nu(\delta a, p_i), \nu(C(\delta a), p_i)) \qquad (1 \leqslant k \leqslant g_i, 1 \leqslant i \leqslant \sigma); \end{aligned} \tag{2.37g}$$

$$\mathfrak{a}^h = \prod_{i=1}^{\sigma} \prod_{k=1}^{g_i} \mathfrak{p}_{ik}^{\nu(\mathfrak{a}^h, \mathfrak{p}_{ik})}, \tag{2.37h}$$

$$\nu(\mathfrak{a}^h, \mathfrak{p}_{ik}) = h\nu(\mathfrak{a}, \mathfrak{p}_{ik}) \qquad (1 \leqslant k \leqslant g_i, 1 \leqslant i \leqslant \sigma, h \in \mathbf{Z}). \tag{2.37i}$$

The numbers a, δ used in (2.37d), (2.37f), (2.37g) refer to a straightened out normal presentation (2.29a) of $\mathfrak{a}$.

The formulae (2.37a–i) freely used certain *divisibility exponents* of the type $\nu(\mathfrak{a}, \mathfrak{b})$ where $\mathfrak{a}$ is any nonzero fractional o_F-ideal and $\mathfrak{b}$ is an ideal of o_F properly contained in o_F.

In general we define $\nu(\mathfrak{a}, \mathfrak{b})$ under the same conditions in the obvious manner as the rational integer satisfying

$$\mathfrak{b}^{\nu(\mathfrak{a},\mathfrak{b})} \mid \mathfrak{a}, \quad \mathfrak{b}^{\nu(\mathfrak{a},\mathfrak{b})+1} \nmid \mathfrak{a} \tag{2.38a}$$

i.e.

$$\mathfrak{b}^{\nu(\mathfrak{a},\mathfrak{b})} \supseteq \mathfrak{a}, \quad \mathfrak{b}^{\nu(\mathfrak{a},\mathfrak{b})+1} \not\supseteq \mathfrak{a}. \tag{2.38b}$$

Note that for the Dedekind domain o_F there holds

$$\bigcap_{\nu=0}^{\infty} \mathfrak{b}^{\nu} = 0 \tag{2.38c}$$

as a consequence of

$$\mathfrak{b} \subset o_F \tag{2.38d}$$

so that $\nu(\mathfrak{a}, \mathfrak{b})$ is uniquely defined.

We have learned already how to calculate $\nu(\mathfrak{a}, \mathfrak{b})$ in case both $\mathfrak{a}$, $\mathfrak{b}$ are normally presented and straightened out.

For any two nonzero ideals $\mathfrak{a}$, $\mathfrak{a}'$ of o_F, we have

$$\nu(\mathfrak{a}\mathfrak{a}', \mathfrak{b}) \geqslant \nu(\mathfrak{a}, \mathfrak{b}) + \nu(\mathfrak{a}', \mathfrak{b}), \tag{2.38e}$$

$$\nu(\mathfrak{a} + \mathfrak{a}', \mathfrak{b}) \geqslant \min(\nu(\mathfrak{a}, \mathfrak{b}), \nu(\mathfrak{a}', \mathfrak{b})), \tag{2.38f}$$

$$\nu(\mathfrak{a} \cap \mathfrak{a}', \mathfrak{b}) \geqslant \max(\nu(\mathfrak{a}, \mathfrak{b}), \nu(\mathfrak{a}', \mathfrak{b})). \tag{2.38g}$$

Furthermore, we have

$$\nu(o_F, \mathfrak{b}) = 0. \tag{2.38h}$$

We define for any nonzero element α of o_F

$$\nu(\alpha, \mathfrak{b}) = \nu(\alpha o_F, \mathfrak{b}), \tag{2.38i}$$

so that the $\mathfrak{b}$-divisibility exponent of α depends only on its equivalence class. Similarly, we define for any nonzero element β of o_F that is not a unit the β-divisibility exponent of the nonzero ideal $\mathfrak{a}$ of o_F according to

$$\nu(\mathfrak{a}, \beta) = \nu(\mathfrak{a}, \beta o_F), \tag{2.38j}$$

so that it only depends on the equivalence class of β.

For any two nonzero ideals $\mathfrak{b}$, $\mathfrak{b}'$ of o_F that are both distinct from o_F, we have

$$\nu(\mathfrak{a}, \mathfrak{b}\mathfrak{b}') \geqslant \min(\nu(\mathfrak{a}, \mathfrak{b}), \nu(\mathfrak{a}, \mathfrak{b}')), \tag{2.38k}$$

$$\nu(\mathfrak{a}, \mathfrak{b} \cap \mathfrak{b}') \leqslant \min(\nu(\mathfrak{a}, \mathfrak{b}), \nu(\mathfrak{a}, \mathfrak{b}')), \tag{2.38l}$$

$$\nu(\mathfrak{a}, \mathfrak{b} + \mathfrak{b}') \geqslant \max(\nu(\mathfrak{a}, \mathfrak{b}), \nu(\mathfrak{a}, \mathfrak{b}')) \quad \text{provided } \mathfrak{b} + \mathfrak{b}' \subset o_F. \tag{2.38m}$$

Finally, let us mention that, for any two rational integers a, b satisfying $a \neq 0$, $b \neq 0, \pm 1$, the divisibility exponent $\nu(a, b)$ is defined by

$$b^{\nu(a,b)} \mid a, \quad b^{\nu(a,b)+1} \nmid a. \tag{2.38n}$$

After the preceding excursion on prime ideal factorizations and divisibility exponents, let us discuss the second way of factoring a given ideal $\mathfrak{a}$ in normal presentation (2.36a) which arises in case a presentation

$$a = b^{\mu} \qquad (\mu \in \mathbf{Z}^{>1}, b \in \mathbf{Z}^{>0}) \tag{2.39a}$$

of a as the μth power of a natural number b is given where $\mu > 1$. In that case we obtain the normal presentation

$$\mathfrak{b} = bo_F + \alpha o_F \tag{2.39b}$$

of an ideal of o_F such that either a nontrivial refinement of the ideal factorization $\mathfrak{a} = \mathfrak{b} \cdot \mathfrak{b}^{-1}\mathfrak{a}$ is possible, or else

$$\mathfrak{a} = \mathfrak{b}^{\mu}, \tag{2.39c}$$

as follows easily by factorization of $\mathfrak{a}$, $\mathfrak{b}$ into products of prime ideal powers. Using the methods of II we succeed in factorizing any finite set of normally presented ideals $\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_m$ of o_F into power products

$$\mathfrak{a}_i = \prod_{j=1}^{t} \prod_{k=1}^{t_j} \mathfrak{a}_{jk}^{\nu(\mathfrak{a}_i, \mathfrak{a}_{jk})} \tag{2.40}$$

of mutually prime ideals of o_F of the normally presented form (2.2a) distinct from o_F such that (2.2b) are natural numbers subject to (2.2c) and that the exponents $\nu(\mathfrak{a}_i, \mathfrak{a}_{jk})$ are nonnegative rational integers.

III. *The Case* $n = 2$. Here we have

$$F = \mathbf{Q}(\sqrt{d}\,), \tag{2.41a}$$

where d is a rational integer such that either $d \equiv 1 \pmod 4$, $d \neq 1$ and d is square free or $d \equiv 0 \pmod 4$, $d \neq 0$ and $d/4$ is square free. In any case

$$o_F = \mathbf{Z}\omega_1 + \mathbf{Z}\omega_2 \qquad \big(\omega_1 = 1, \omega_2 = (d + \sqrt{d}\,)/2\big), \tag{2.41b}$$

and the rational content of

$$\gamma = e_1\omega_1 + e_2\omega_2 \qquad (e_1, e_2 \in \mathbf{Z})$$

is given by

$$C(\gamma) := \gcd(e_1, e_2). \tag{2.41c}$$

The automorphism group of F is of order 2 with

$$\sigma\colon F \to F,\quad \sigma(e_1\omega_1 + e_2\omega_2) = (e_1 + e_2 d)\omega_1 - e_2\omega_2 \qquad (e_1, e_2 \in \mathbf{Q}) \tag{2.41d}$$

the generating automorphism. For any normal presentation (2.25a) of a fractional o_F-ideal $\mathfrak{a}$ of o_F, one obtains the normal presentation

$$\sigma(\mathfrak{a}) = \delta(ao_F + \sigma(\alpha)o_F)$$

of the *conjugate ideal* $\sigma(\mathfrak{a})$. For the norm of $\mathfrak{a}$ one obtains the normal presentation

$$N(\mathfrak{a})o_F = \mathfrak{a}\sigma(\mathfrak{a}) = \delta^2(a^2 o_F + N(\alpha)o_F),$$

where $N(\alpha) = \alpha\sigma(\alpha) \in \mathbf{Q}$; in other words

$$N(\mathfrak{a}) = \delta^2 \gcd(a^2, N(\alpha)). \tag{2.41e}$$

Hence

$$\mathfrak{a}^{-1} = N(\mathfrak{a})^{-1}\sigma(\mathfrak{a}), \tag{2.41f}$$

in case $\mathfrak{a} \neq 0$. Due to (2.41e, f) the content calculus becomes particularly simple for quadratic number fields. It is desirable only to employ ideals $\mathfrak{a}$ of o_F with normal presentation (2.36a) such that either

$$a \mid d, \quad \text{hence } \sigma(\mathfrak{a}) = \mathfrak{a} \tag{2.41g}$$

or

$$\gcd(a, d) = 1, \quad \text{hence } \mathfrak{a} + \sigma(\mathfrak{a}) = o_F. \tag{2.41h}$$

3. Index Estimates. Let us assume that we have already determined units $\varepsilon_1, \ldots, \varepsilon_r$ generating a subgroup of finite index in the full unit group U_F of F. By TU_F we denote the torsion subgroup of U_F. We set $U_\varepsilon := TU_F \times \langle \varepsilon_1, \ldots, \varepsilon_r \rangle$. According to [3] we can assume that U_ε contains all units of U_F which already lie in proper subfields of F.

In this section we will give an upper bound for $(U_F : U_\varepsilon)$. This is done usually by means of a lower bound for the regulator R_F of F using the relation

$$(U_F\colon U_\varepsilon) = \frac{R_\varepsilon}{R_F}, \tag{3.1}$$

where R_ε denotes the regulator computable from the independent units $\varepsilon_1, \ldots, \varepsilon_r$ of U_ε by

$$R_\varepsilon := \operatorname{abs}\left(\det\left(c_j \log|\varepsilon_i^{(j)}|\right)\right) \qquad \left(i, j = 1, \ldots, r;\ c_j = \begin{cases} 1 & \text{for } j \leq r_1 \\ 2 & \text{else} \end{cases}\right). \tag{3.2}$$

Lower bounds for R_F are given in [5], [2]. The first type of estimates depends only on the field degree n but is very good for small values of $|d_F|$. Those estimates were derived by analytic methods. The second type of estimates depends on n and $|d_F|$. They are, generally speaking, not as sharp; they were obtained by number geometric methods. In [2] only totally real number fields were considered. Fortunately by the same arguments analogous results are obtained for all signatures.

THEOREM (3.1). *If F is primitive, the regulator R_F of F satisfies the lower estimate*

$$R_F \geq \left[\left(\frac{12 \log^2 \sqrt{|d_F|/n^n}}{(n-1)n(n+1) - 6r_2}\right)^r \frac{2^{r_2}}{\gamma_r^r n}\right]^{1/2}. \tag{3.3}$$

Here and in the following γ_r^r denotes Hermite's constant for positive definite quadratic forms. This result can still be improved by solving a minimization problem analogous to (B) in [2]. For imprimitive fields one proceeds as in Satz XII of [2].

In case these estimates are not good enough for our purposes, we derive a new lower bound for R_F by using successive minima of the integers of F.

The mapping

$$\psi\colon o_F \to \mathbf{C}^n\colon \omega \to \left(\omega^{(1)}, \ldots, \omega^{(n)}\right) \tag{3.4}$$

provides a $\mathbf{Z}$-lattice in the complex space $\mathbf{C}^n$ with basis $\psi(\omega_1), \ldots, \psi(\omega_n)$, where $\omega_1, \ldots, \omega_n$ is an arbitrary integral basis of F. By decomposing the complex conjugates into their real and imaginary parts we can also consider this lattice as a subset of the n-dimensional Euclidean space $\mathbf{R}^n$. On $\psi(o_F)$ we introduce the distance function

$$\|\psi(\omega)\| := \left(\sum_{j=1}^{n} |\omega^{(j)}|^2\right)^{1/2} = \left(\psi(\omega)\,\overline{\psi(\omega)}^{\,t}\right)^{1/2}. \tag{3.5}$$

According to Minkowski we define n successive minima on $\psi(o_F)$.

$$\begin{aligned} M_1 &:= \min\{\|x\| \mid x \in \psi(o_F), x \neq 0\} = \|Y_1\| \quad \text{and} \\ M_{k+1} &:= \min\{\|x\| \mid x \in \psi(o_F);\ x, Y_1, \ldots, Y_k \text{ lin.ind.}\} = \|Y_{k+1}\| \\ &\qquad\qquad (k = 1, \ldots, n-1) \end{aligned} \tag{3.6}$$

with the properties

$$M_1 \leqslant M_2 \leqslant \cdots \leqslant M_n \tag{3.7}$$

and

$$M_1 \cdot \cdots \cdot M_n \leqslant \sqrt{\gamma_n^n |d_F|}\,. \tag{3.8}$$

Obviously, we can assume $M_1 = \sqrt{n}$, $Y_1 = \psi(1)$, since every nonzero lattice vector has length at least $\sqrt{n}$ according to the inequality between arithmetic and geometric means and the fact that the product of its coordinates is the norm of an integer of F and therefore of absolute value at least one. This argument also proves

$$\|\psi(\xi)\| = \sqrt{n} \Leftrightarrow \xi \in TU_F. \tag{3.9}$$

For $\varepsilon \in U_F \setminus U_\varepsilon$ any n consecutive powers, for example, $\varepsilon^{1+[n/2]-n}, \ldots, \varepsilon^{-1}$, $\varepsilon^0, \varepsilon^1, \ldots, \varepsilon^{[n/2]}$ are linearly independent over $\mathbf{Q}$. Hence the vectors

$$\psi(\varepsilon^\nu) \qquad (\nu = 1 + [n/2] - n, \ldots, 0, \ldots, [n/2])$$

are linearly independent in $\psi(o_F)$. We order them according to the size of their lengths and obtain a bijection $\pi\colon \{1, \ldots, n\} \Leftrightarrow \{1 + [n/2] - n, \ldots, [n/2]\}$ such that $\|\psi(\varepsilon^{\pi(1)})\| \leqslant \cdots \leqslant \|\psi(\varepsilon^{\pi(n)})\|$. Again we can choose $\pi(1) = 0$. From the definition of the successive minima we know that

$$M_i \leqslant \|\psi(\varepsilon^{\pi(i)})\| \qquad (i = 1, \ldots, n).$$

LEMMA (3.1). $\|\psi(\varepsilon^k)\| \leqslant \|\psi(\varepsilon^{k+1})\|$ *for all* $k \in \mathbf{N}$.

This is easily seen by the Lagrange multiplier method using $|N(\varepsilon)| = 1$.

Since we can consider ε^{-1} instead of ε, the worst possible estimates occur for $\|\psi(\varepsilon^0)\| \leqslant \|\psi(\varepsilon)\| \leqslant \|\psi(\varepsilon^{-1})\| \leqslant \ldots$; that is, we obtain

$$M_{2k+1} \leqslant \max\{\|\psi(\varepsilon^k)\|, \|\psi(\varepsilon^{-k})\|\} \quad \text{for } k = 1, 2, \ldots, \left[\frac{n}{2}\right].$$

Hence, for every $\varepsilon \in U_F \setminus U_\varepsilon$ or its inverse, we have

$$\|\psi(\varepsilon)\| \geqslant \max\left\{\left\{M_{2k+1}^{1/k} \mid k = 1, \ldots, [n/2] - 1\right\} \cup \left\{M_n^{1/[n/2]}\right\}\right\} =: M. \tag{3.10}$$

For the next lemma we would like to have $\|\psi(\varepsilon)\| > \sqrt{n}$. Unfortunately this can be guaranteed only in case the successive minima of the lattice are not obtained from torsion elements of U_F. If the latter is the case, we must use the results from [5], [2] at the beginning of this section.

LEMMA (3.2). *In case $M^2 > n$, the minimum of $\Sigma_{j=1}^n (\log|\varepsilon^{(j)}|)^2$ for $\varepsilon \in U_F \setminus U_\varepsilon$ is at least*

$$n\left[\log\left(\frac{M^2}{n} + \left(\frac{M^4}{n^2} - 1\right)^{1/2}\right)^{1/2}\right]^2 = m.$$

Proof. We set $x_j := |\varepsilon^{(j)}|$ and apply the Lagrange multiplier method to

$$\sum_{j=1}^{n} (\log x_j)^2 \quad \text{subject to} \quad \sum_{j=1}^{n} \log x_j = 0 \text{ and } \sum_{j=1}^{n} x_j^2 \geqslant M^2.$$

The side conditions correspond to $|N(\varepsilon)| = 1$ and (3.10), where we need not make a difference between ε and ε^{-1} any longer. (Obviously we must have equality in the second side condition for a minimum.) It is easily seen that there are at most two different values, say $x_1 = \cdots = x_s = x$ and $x_{s+1} = \cdots = x_n = y$ for an optimal solution. Another extremal value problem yields $s = n/2$** and thus the result of the lemma.

The index estimates we are looking for can now be easily derived as in [3].

Namely, representing ε by a system of r fundamental units, $\Sigma_{j=1}^n (\log|\varepsilon^{(j)}|)^2$ turns out to be a positive definite quadratic form in r variables. Its determinant is easily calculated as $2^{-r_2} n R_F^2$, and Minkowski's theorem on successive minima yields

$$m^r \leqslant \gamma_r^r 2^{-r_2} n R_F^2 \tag{3.11}$$

hence a lower bound for the regulator of the field from which an upper bound for the index $(U_F : U_\varepsilon)$ follows. In case intermediary subfields occur this bound must be slightly changed as was outlined already in [3].

Mathematisches Institut
Universität Düsseldorf
Universitätsstr. 1
4 Düsseldorf, West Germany

Department of Mathematics
The Ohio State University
Columbus, Ohio 43210

** This was suggested to the first author by H. W. Lenstra, Jr., abridging lengthy computations which were needed in [2] for a slightly stronger result.

1. R. KANNAN, *A Polynomial Algorithm for the Two-Variable Integer Programming Problem*, Report No. 78116, Institut für Ökonometrie und Operations Research, Universität Bonn.

2. M. POHST, "Regulatorabschätzungen für total reelle algebraische Zahlkörper," *J. Number Theory*, v. 9, 1977, pp. 459–492.

3. M. POHST & H. ZASSENHAUS, "An effective number geometric method of computing the fundamental units of an algebraic number field," *Math. Comp.*, v. 30, 1977, pp. 754–770.

4. M. POHST & H. ZASSENHAUS, "On unit computation in real quadratic fields," in *Symbolic and Algebraic Computation*, Lecture Notes in Comput. Sci., v. 72, 1979, pp. 140–152.

5. R. ZIMMERT, *Ideale kleiner Norm in Idealklassen und eine Regulatorabschätzung*, Thesis, Bielefeld, 1978.