Non-homogeneous Binary Quadratic Forms

K. Inkeri

This is a paper of Inkeri, re-typed by myself, with some slight changes in the original notation. The article was first published in 1949 in Skandinaviske Matematikerkongress, Trodheim, pp. 216 – 224.

Let $\alpha, \beta, \gamma, \delta$ be real numbers with $\Delta = \alpha\delta - \beta\gamma \neq 0$. A classical theorem of Minkowski states that for any real numbers λ, μ there exist integers x, y such that

$$|(\alpha x + \beta y - \lambda)(\gamma x + \delta y - \mu)| \leq \frac{1}{4}|\Delta|. \tag{1}$$

Here the product $(\alpha x + \beta y)(\gamma x + \delta y)$ is an indefinite binary quadratic form with discriminant Δ^2. Consequently, we may express Minkowski's theorem also in the form: if

$$f(x, y) = ax^2 + bxy + cy^2$$

is any indefinite binary quadratic form with real coefficients a, b, c and discriminant $d = b^2 - 4ac$, then for any real x_0, y_0 there exist real x, y with

$$x \equiv x_0 \pmod 1, \; y \equiv y_0 \pmod 1 \tag{2}$$

such that

$$|f(x, y)| \leq \frac{1}{4}\sqrt{d}. \tag{3}$$

Many proofs have been given for this theorem, and the most recent of these leave nothing to be desired in brevity. Lately many attempts have also been made to generalize and improve the result. In particular, certain papers of Heinhold[1], Davenport[2], Varnavides[3] and Cassels[4] should be mentioned. It is well known that the constant on the right hand side of (3) is the best possible only for a class of forms of minor importance. Attempts to improve the result have been based mostly on the following definition:

Let $M(f)$ (briefly M) denote the lower bound of all the numbers m for each of which there exist for any x_0, y_0 numbers x, y satisfying (2) such that

$$|f(x, y)| \leq m. \tag{4}$$

If now also M belongs to the set of m's then we denote it, following Heinhold (H.I, p. 660) by M_1, otherwise by M_2. In the case $M = M_1$, the inequality

$$|f(x, y)| \leq M$$

has therefore always a solution in the above sense.

If, on the other hand, $M = M_2$, (4) does not have a solution for all x_0, y_0 and conversely. Plainly, equivalent forms have the same lower bound M, which therefore is a constant, characteristic of the whole form class.

[1] J. Heinhold, Math. Z., 44 (1939), 659 – 688 and 45 (1939), 176–184. We refer to these two papers as H.I and H.II.

[2] H. Davenport, Proc. Kon. Ned. Akad. Wet. (Amsterdam), 49 (1946), 815–821 (D.I); 50 (1947), 378–389 (D.II); 484–491(D.III); 741–749, and 909–917 (D.IV).

[3] P. Varnavides, Quart. J. Math. (Oxford), 19 (1948), 54–58 (V.I), Proc. Kon. Ned. Akad. Wet. (Amsterdam), 51 (1948), 396–404, and 470–481 (V.II)

[4] J.W.S. Cassels, Proc. Cambridge Phil. Soc., 44 (1948), 145–154 (C.I) and 457–462 (C.II).

Corresponding constants M may also be defined for definite and semi-definite forms. According to the investigations of Dirichlet (Werke II, pp. 29–48) and Heinhold (H.I, 661-667), M may be given in terms of the coefficients of the appropriate equivalent form. But, on the other hand, it seems unreasonable to expect that M can be expressed in a corresponding manner for indefinite forms, even though it has been possible to determine exact values of M for a relatively large number of such forms. Hitherto the object of the studies has been primarily to find good upper and lower bounds for M. From Minkowski's result it follows that in all cases

$$M \leq \frac{1}{4}\sqrt{d}.$$

The sign of equality is necessary only for the forms that are equivalent to the form bxy. For other forms, Heinhold (H.I cf. especially p. 675) has been able to replace the righthand member by a smaller expression in terms of coefficients of the form. However, Davenport's results (D.I p. 816) concerning forms that do not represent zero is much more simple. In this result, an essential part of which Varnavides (V.I p. 55) has improved somewhat, the upper bound obtained may be expressed as a function of the discriminant and the first coefficient of an appropriate equivalent form. By a more detailed study I have been able to improve Davenport's theorem. Application of the results obtained has in certain special cases furnished good estimates for M, and in some cases that previously required complicated special treatment, even precise values for M.[5]

As regards lower bounds for M, we must above all call attention to the fact that Davenport[6] succeeded last year in proving the following very remarkable result:

$$M \geq \frac{1}{128}\sqrt{d}.$$

This holds for any indefinite binary quadratic form that does not represent zero. Our question concerning M is closely connected with the problem of the validity of Euclid's algorithm in real quadratic fields. Let d be the discriminant of the real quadratic field $\mathbb{Q}(\sqrt{d})$, and let

$$f_d(x,y) = \begin{cases} x^2 - \frac{d}{4}y^2 & d \equiv 0 \pmod 4 \\ x^2 + xy + \frac{1-d}{4}y^2 & d \equiv 1 \pmod 4 \end{cases}$$

so that $f_d(x,y)$ is the principal form belonging to the discriminant d. Euclid's algorithm is valid in $\mathbb{Q}(\sqrt{d})$ if and only if there exist for any rational x_0, y_0 numbers x, y satisfying (2) such that $|f_d(x,y)| < 1$. If, in the above definition of M, we restrict ourselves only to rational numbers x_0, y_0, then we obtain for the m's, instead of M, a new upper bound, say M_0. It is plain that $M_0 \leq M$. If $M_0 < 1$ (which is certainly so if $M < 1$), then Euclid's algorithm holds in $\mathbb{Q}(\sqrt{d})$, but it does not hold if $M_0 > 1$. We will pass over the case $M_0 = 1$ which also may occur (viz. for $d = 65$). For M_0 one gets again Davenport's lower bound $2^{-7}\sqrt{d}$.[7]

Since Euclid's algorithm does not hold in $\mathbb{Q}(\sqrt{d})$ when $d > 100$ (cf. Hua's review of my paper in Math. Rev. 10 (1949), p. 15), we infer that $M_0 \geq 1$, and hence also $M \geq 1$, which for small d (viz. for $100 < d < 2^{14}$) is better than Davenport's result.

In the following we shall discuss in greater detail a special question concerning M. As we have mentioned above, the precise values of M are known in many cases (cf. H.I, pp. 675 – 680, 683–684; H.II; D.I, p. 3; D.III; V.I, p. 58; C.II, p. 462). As a simple example the following result of mine will serve: let

$$\frac{d}{4} = q^2 + p = (q+1)^2 - p_1$$

[5] These results will be presented elsewhere.

[6] This work has not, as far as I know, yet appeared from the press (cf. the abstract in Bull. Amer. Math. Soc. 54 (1948), p. 662).

[7] This result of Davenport has been conclusive for the question relating to Euclid's algorithm. Using this and earlier results (cf. Inkeri, Ann. Acad. Sci. Fenn., 1947, A, I, N:o 41), I have been able to bring this problem to an end (i.e. to treat the remaining cases: d is a prime with $5000 < d < 2^{14}$ and $d \equiv 1 \pmod{24}$). I informed Prof. Davenport of these results in September, 1948. Chatland (cf. the abstract in Bull. Amer. Math. 54 (1948), p. 828) has independently obtained the same results.

where p, p_1 and q are positive integers (and therefore $q = \left\lfloor \frac{1}{2}\sqrt{d} \right\rfloor$). If either q is divisible by p or $q+1$ is divisible by p_1 then

$$M(f_d) = \frac{1}{4}\max\{p, p_1\}.$$

It is worth remarking in this connection that in every case up to now in which the exact value of M is known is for some indefinite form, it happens that $M = M_1$. According to Heinhold (H.I, p. 662) the possibility $M = M_2 = 0$ occurs for semi-definite forms fairly often. We shall now show, by an example, that there also exist indefinite forms, even indefinite forms with integral coefficients, for which $M = M_2$.

We consider the principal form of the discriminant $d = 13$, i.e.

$$f_{13}(x, y) = x^2 + xy - 3y^2 = (x + \omega y)(x + \overline{\omega} y),$$

where

$$\omega = \frac{1}{2}(1 + \sqrt{13}) = 2.302..., \; \overline{\omega} = \frac{1}{2}(1 - \sqrt{13}) = -1.302...$$

and hence

$$\omega + \overline{\omega} = 1, \; \omega\overline{\omega} = -3.$$

The set $\{1, \omega\}$ forms a basis of the quadratic field $K = \mathbb{Q}(\sqrt{13})$, and f_{13} is also the norm of the general integer $x + y\omega$ in K.

To prove that in fact $M(f_{13})$ has the property $M = M_2$, we show in the first place that $M \leq \frac{1}{3}$. This result can be reached by two essentially different methods, one due to Davenport (cf. D.II) and the other my own.[8] We shall here employ the former, as it is more closely connected with the procedure that leads to the final relation $M = M_2$.

Following the first presentation of Minkowski's theorem we define $M(\lambda, \mu)$ as the lower bound of all numbers

$$|(\xi - \lambda)(\overline{\xi} - \mu)| \tag{5}$$

where ξ runs through all integers of the field K.

To prove $M \leq \frac{1}{3}$, it evidently suffices if we show that for any λ, μ, $M(\lambda, \mu) \leq \frac{1}{3}$. We shall consider arbitrary numbers λ, μ such that $M(\lambda, \mu) \neq 0$.

Let ϵ_0 denote a sufficiently small positive number which is smaller than any of a certain finite number of absolute positive constants, which will be defined more accurately (in general, implicitly) in the course of our proof.

According to Davenport (D.II, p. 380) there exist, for real λ, μ, real α, β, ϵ such that

$$M(\lambda, \mu) = \frac{1 - \epsilon}{|\alpha\beta|}, \; 0 \leq \epsilon \leq \epsilon_0; \tag{6}$$

$$\alpha \geq \sqrt{2} - \epsilon > 0, \; \frac{\alpha}{\eta} \leq |\beta| \leq \alpha \tag{7}$$

and furthermore such that

$$P(\xi) = |(\alpha\xi - 1)(\beta\overline{\xi} - 1)| \geq 1 - \epsilon \tag{8}$$

for all $\xi \in \mathcal{O}_K$. Here η denotes the fundamental unit of the field K, whence

$$\eta = \frac{1}{2}(3 + \sqrt{13}) = \omega + 1, \; N(\eta) = 1.$$

By (6) we need only show that

$$|\alpha\beta| \geq 3(1 - \epsilon). \tag{9}$$

[8] I am using this method in other studies, and will give an account of it in another paper.

If $\beta > 0$, this may be verified easily as follows. In the first place $\beta > 1$, for otherwise by (7),

$$P(1) = (\alpha - 1)(1 - \beta) \leq \eta\beta(1 - \beta) \leq \eta \left(\frac{\beta + (1 - \beta)}{2}\right)^2 = \frac{\eta}{4} < 1,$$

which is inconsistent with (8) (if ϵ_0 sufficiently small). On the other hand, by using (8) and taking $\xi = 1$,

$$\left(\frac{\alpha + \beta}{2} - 1\right)^2 \geq (\alpha - 1)(\beta - 1) = \alpha\beta - (\alpha + \beta) + 1 \geq 1 - \epsilon,$$

whence

$$\alpha\beta \geq (\alpha + \beta) - \epsilon \geq 2(\sqrt{1 - \epsilon} + 1) - \epsilon \geq 4 - 3\epsilon.$$

This implies even more than what was required by (9).

Hence we may write $-\beta = \beta_0 > 0$. Now we shall prove that $\beta_0 \leq 1$ is impossible. The proof is based on the three inequalities derived from (8) by taking ξ equal to -1, ω and -2.

Since $\alpha \leq \eta\beta_0$ and therefore $P(-1) = (\alpha + 1)(1 - \beta_0) \leq (1 + \eta\beta_0)(1 - \beta_0)$ we have, by the inequality[9] $P(-1) \geqslant 1$,

$$\eta\beta_0 - \eta + 1 \leqslant 0$$

and further

$$\alpha \leq \eta\beta \leqslant \omega, \quad \beta_0 \leqslant \frac{\omega}{\eta} < \frac{\omega}{3}. \tag{10}$$

By the second half of (10), $P(\omega) = (3\alpha + \overline{\omega})\left(\frac{\omega}{3} - \beta_0\right) \geqslant 1$, both the factors on the left being positive. From this, on noting that $3\alpha + \overline{\omega} \leqslant 3\omega + \overline{\omega} = 2\omega + 1$, it follows that

$$\beta_0 \leqslant \frac{\omega}{3} - \frac{1}{2\omega + 1} = \frac{\omega + 3}{9} < \frac{3}{5}, \tag{11}$$

and hence $\alpha < \frac{3}{5}\eta < 2$.

Now the relation $P(-2) \geqslant 1$ gives

$$|2\beta_0 - 1| > \frac{1}{5}.$$

But from the conditions (7) we obtain

$$\beta_0 \geq \frac{\alpha}{\eta} > \frac{7}{5}|\overline{\eta}| > \frac{2}{5}$$

which, together with (11) contradicts the last inequality. This completes the proof that $\beta_0 > 1$, i.e. $\beta < -1$.

We obtain from $P(-1) \geq 1 - \epsilon$

$$\beta_0 \geq 1 + \frac{1}{\alpha + 1} - \frac{\epsilon}{\alpha + 1},$$

so that, for $\alpha \geq \omega$,

$$\alpha\beta_0 \geq \alpha + 1 - \frac{1}{\alpha + 1} - \frac{\alpha}{\alpha + 1}\epsilon \geq \omega + 1 - \frac{1}{\omega + 1} - \epsilon = \eta + \overline{\eta} - \epsilon$$

and hence (9) is valid. We may therefore restrict ourselves to consider only the possibility $\alpha < \omega$.

Suppose first that $\beta_0 \geq \frac{1}{4} - \overline{\omega} = 1.55....$ It is plain that (9) then holds for $\alpha \geq 2$. For $\alpha < 2$, the relation $P(2\eta) \geqslant 1$ gives

$$2\left|\beta + \frac{\eta}{2}\right| \geq \left|\alpha + \frac{\overline{\eta}}{2}\right| \cdot \left|\beta + \frac{\eta}{2}\right| \geqslant \frac{1}{4},$$

[9] The nature of this argument in the following will in general be such that we can neglect ϵ. For clarity the signs $\geq$ and $\leq$ will be replaced by $\geqslant$ and $\leqslant$, respectively, to indicate that to the expressions following these signs must be added respectively a term $-C\epsilon$ or $C\epsilon$ (C a positive absolute constant) until either a "$>$" or a "$<$" is reached.

from which it follows that

$$\beta_0 \geqslant \frac{\eta}{2} + \frac{1}{8} > \sqrt{3},$$

since $\frac{\eta}{2} - \beta_0 < 1.65... - 1.55 < \frac{1}{8}$. Consequently, $\alpha\beta_0 \geq \beta_0^2 > 3$ by (7).

Suppose now that $\beta_0 < \frac{1}{4} - \overline{\omega}$. Since $\beta < -1$ and $\alpha \geqslant \sqrt{2}$, we may write

$$\alpha = \omega - \delta, \quad 0 < \delta < 1$$

$$\beta = \overline{\omega} - \theta, \quad 0 < \theta < \frac{1}{4}. \tag{12}$$

Substituting these values of α and β in (8) and taking ξ to be equal to

$$\xi := \frac{1-\eta^n}{\omega} = -\frac{1-\eta^n}{1-\eta}, \tag{13}$$

which evidently is an integer of K for every rational integer n, we obtain, after a simple reduction

$$|(1-\lambda_n\delta)(1-\mu_n\theta)| \geq 1-\epsilon \tag{14}$$

where

$$\lambda_n = \frac{1-\eta^{-n}}{\omega},\ \mu_n = \overline{\lambda_n} = \frac{1-(-\eta)^n}{\overline{\omega}}.$$

The inequality (13) is valid for any rational integer n, but we shall employ it only when n is nonnegative.

We must now prove that $\theta = 0$. Since λ_n increases as n increases and

$$\lim_{n\to\infty} \lambda_n = \frac{1}{\omega}, \tag{15}$$

we have for $n \geq 0$,

$$1 = 1-\lambda_0\delta \geq 1-\lambda_n\delta > 1-\frac{1}{\omega} > 0, \tag{16}$$

so that $|1-\lambda_n\delta| \leq 1$. Hence we obtain from (14) that

$$|1-\mu_n\theta| \geq 1-\epsilon, \quad n \geq 0. \tag{14*}$$

If initially $\theta < 0$, we choose a positive integer m such that

$$|\mu_{2m-1}\theta| \leq 1 < |\mu_{2m+1}\theta|,$$

which is possible since $|\mu_1\theta| = \eta|\theta| < 1$, by (12), and since $|\mu_{2m+1}|$ tends to ∞ as m tends to ∞. Noting that $\mu_{2m-1} < 0$ and

$$\left|\frac{\mu_{2m+1}}{\mu_{2m-1}}\right| = \frac{1+\eta^{2m+1}}{1+\eta^{2m-1}} < \eta^2,$$

it follows from these inequalities that

$$0 \leq 1-|\mu_{2m-1}\theta| < 1-\frac{|\mu_{2m+1}\theta|}{\eta^2} < 1-\eta^{-2},$$

which contradicts (14*); hence it is not true that $\theta < 0$. If, again, $\theta > 0$, then by $\mu_2 = (1-\eta)\mu_1 = 1+2\eta$ and (12),

$$1-\mu_2\theta > \frac{3-2\eta}{4} > -(1-\epsilon).$$

Now (14*) gives $1-\mu_2\theta \geq 1-\epsilon$, whence $\mu_2\theta \leq \epsilon < 1$. Since furthermore, $\mu_{2m} > 0$ for $m > 0$ and μ_{2m} tends to ∞ as m tends to ∞, there exists a positive integer m such that

$$\mu_{2m}\theta \leq 1 < \mu_{2m+2}\theta.$$

By $\frac{\mu_{2m+2}}{\mu_{2m}} \leq \eta^2 + 1$, we obtain

$$0 \leq 1 - \mu_{2m}\theta \leq 1 - \frac{\mu_{2m+2}}{\eta^2+1} < 1 - \frac{1}{\eta^2+1}.$$

This contradiction proves $\theta > 0$ is impossible. Consequently, $\theta = 0$ and (14) becomes

$$|1 - \lambda_n \delta| \geq 1 - \epsilon.$$

By (16) this gives $\lambda_n \delta \leq \epsilon$ for every positive integer n. Using (15) we deduce that also

$$\frac{1}{\omega}\delta \leq \epsilon, \qquad \text{i.e. } \alpha \geq \omega(1-\epsilon).$$

Therefore, $|\alpha\beta| \geq |\omega(1-\epsilon)\overline{\omega}| = 3(1-\epsilon)$, and thus (9) is valid. This completes the proof that $M \leq \frac{1}{3}$.

We now proceed to prove that $M = M_2$. At the same time it will be found that $M = \frac{1}{3}$. This latter result, however, may also be obtained directly by showing that

$$\left| N\left(\xi - \frac{\omega}{3}\right) \right| = \left| \left(\xi - \frac{\omega}{3}\right)\left(\overline{\xi} - \frac{\overline{\omega}}{3}\right) \right| \geq \frac{1}{3} \tag{17}$$

for every $\xi \in \mathcal{O}_K$, i.e. $M\left(\frac{\omega}{3}, \frac{\overline{\omega}}{3}\right) = \frac{1}{3}$. Let us put $\xi = \frac{1}{2}(r + s\sqrt{13})$, where r and s are rational integers and $r \equiv s \pmod 2$. Then

$$N\left(\xi - \frac{\omega}{3}\right) = \frac{1}{36}\left[(3r-1)^2 - 13(3s-1)^2\right],$$

where the expression in the square brackets is congruent to 0 both mod 3 and mod 4, and therefore also mod 12. Consequently, this expression is numerically ≥ 12, from which (17) immediately follows.

To prove that $M = M_2$, it suffices if we show that there exist λ, μ such that the expression (5) is $> \frac{1}{3}$ for any $\xi \in \mathcal{O}_K$. Numbers having this property are $\lambda = \frac{\omega}{3}$ and $\mu = -\frac{\omega}{3}$. To establish this, let us suppose that the integer ξ satisfies the condition

$$P = |3\xi - \omega| \cdot |3\overline{\xi} + \omega| \leq 3. \tag{18}$$

First we may infer $\xi > 0$. In fact, if $\xi \leq 0$, then we would have

$$|3\xi - \omega| = |3\xi + \overline{\omega} - 1| = |3\xi - \overline{\omega}| + 1$$

since $\overline{\omega} < 0$ and therefore

$$P = |N(3\overline{\xi} + \omega)| + |3\overline{\xi} + \omega|. \tag{18*}$$

But now $3\overline{\xi} - \omega \neq 0$, for if $3\overline{\xi} = -\omega$, it would follow because $\omega\overline{\omega} = -3$, that $9N(\xi) = -3$, which is impossible. Furthermore, $|N(3\overline{\xi} + \omega)|$ is also $\neq 0$ and divisible by 3, and hence ≥ 3. Therefore, by (18*), $P > 3$, contrary to (18).

Next, we show the condition

$$|N(3\overline{\xi} + \omega)| = 3$$

leads to a contradiction. From the condition it follows at once that

$$|N(\omega\xi - 1)| = 1,$$

i.e. $\omega\xi - 1$ is a unit in K. Since η was the fundamental unit of K, we may write

$$\omega\xi - 1 = \pm\eta^n,$$

where n is a rational integer. But now $\eta \equiv 1 \pmod{\omega}$ and hence only the minus sign on the right hand side is possible. Thus

$$\xi = \frac{1 - \eta^n}{\omega},\ \overline{\xi} = \frac{1 - (-\eta)^{-n}}{\overline{\omega}},$$

whence

$$P = \omega\eta^{-n} + 3 > 3$$

contrary to (18) as required.

Since 6 and -6 are not norms in K, we must have

$$|N(3\overline{\xi} + \omega)| \geq 9.$$

The inequality (18) and the latter give on division

$$\left|\frac{3\xi - \omega}{3\xi + \overline{\omega}}\right| \leq \frac{1}{3}.$$

If $\xi > \frac{\omega}{3}$, then it follows further, by $\omega > |\overline{\omega}|$,

$$\xi \leq \frac{2 - \sqrt{13}}{6} < 1.$$

But $\xi > 0$, as proved above, and $\frac{\omega}{3} < 1$, whence in every case $|\xi| < 1$.

Noting that $-\xi = \overline{-\overline{\xi}}$, (18) may be also written in the form

$$|3(-\overline{\xi}) - \omega| \cdot |3\left(\overline{-\overline{\xi}}\right) + \omega| \leq 3.$$

Consequently, in addition to ξ, also $-\overline{\xi}$ satisfies the condition (18) and we deduce, as for ξ, that $|-\overline{\xi}| < 1$. But then $|N(\xi)| < 1$, which is impossible, since $\xi \in \mathcal{O}_K$, and not zero. This completes our proof for $M = M_2$.

Finally we may mention that $M(\lambda, \mu)$ is equal to $\frac{1}{3}$ for (5) only if λ, μ belongs to a comparatively restricted set of real numbers. By using the methods applied in this paper it is also possible to show that $M(\lambda, \mu) \leq \frac{4}{13}$ when λ, μ are such that $M(\lambda, \mu) \neq \frac{1}{3}$.