ON THE MINIMUM OF A TERNARY CUBIC FORM

WERNER GEORG NOWAK
Institut für Mathematik der Universität für Bodenkultur
Gregor Mendel-Strasse 33, A-1180 Vienna, Austria
(Submitted May 1984)

Let

$$f_a = f_a(x, y, z) = x^3 + y^3 + z^3 + 3axyz, \tag{1}$$

a an arbitrary real constant, and denote, for a lattice Λ in $\mathbb{R}^3$, by $\mu_a(\Lambda)$ the infimum of $|f_a|$ if (x, y, z) runs through all lattice points of Λ except (0, 0, 0). It is the objective of the present paper to estimate, from the above, the supremum M_a of $\mu_a(\Lambda)$, taken over all lattices Λ with lattice constant 1. (Since any homogeneous ternary cubic polynomial can be transformed into the shape (1) by a suitable linear transformation, there is no loss of generality in starting from this canonical form.)

Classical work on this topic has been done by Mordell [6] (on the basis of his method of reducing the problem to a two-dimensional one) and by Davenport [1], [2]. Significant progress has been achieved in the special case $a = 0$. For arbitrary a however, the results obtained were not very sharp, as was noted by Golser [3], who improved upon Mordell's estimate for the general case, by a refined variant of his method. Later on, in [4], he observed that, for a certain range of the constant a, the bound can be improved further by the simple idea of inscribing a sphere into the star body $|f_a| \leqslant 1$.

The purpose of this short note is to establish a result that improves upon all known estimates for certain intervals of a (at least for $0.9 \leqslant a \leqslant 2.9$ and for $-6 \leqslant a \leqslant -1.2$; see the tables at the end) by the elementary procedure of inscribing an ellipsoid of the shape

$$E_t(r) : x^2 + y^2 + z^2 + 2t(xy + xz + yz) \leqslant r^2, \tag{2}$$

where t is a parameter with $-\frac{1}{2} < t < 1$, into the body $K_a: |f_a| \leqslant 1$. Our result reads

Theorem: For arbitrary real a and a parameter t with $-\frac{1}{2} < t < 1$, $t \neq 0$, we have

$$M_a \leqslant \sqrt{2}(1 - t)\sqrt{1 + 2t}\, m_a(t),$$

where

$$m_a(t) := \max\{|1 + a|(1 + 2t)^{-3/2}3^{-1/2},\ \phi_1(t),\ \phi_2(t)\},$$
$$\phi_j(t) := (2 + 2t + 4tc_j + c_j^2)^{-3/2}\,|2 + 3ac_j + c_j^3| \qquad (j = 1, 2),$$
$$c_j := (2t)^{-1}(b - 2t + (-1)^j(b^2 + 4t + 4bt^2)^{1/2}),\ b = a - 1.$$

Proof: We first briefly recall some well-known facts from the *Geometry of Numbers*. The critical determinant $\Delta(K_a)$ of our body K_a is defined as the infimum of all lattice constants $d(\Lambda)$ of lattices Λ in $\mathbb{R}^3$ which have no point in the interior of K_a except the origin. For any such lattice Λ, we put

$$\Lambda_1 = d(\Lambda)^{-1/3}\Lambda,$$

[such that $d(\Lambda_1) = 1$] and $\Lambda' = \Lambda - \{(0, 0, 0)\}$, $\Lambda_1' = \Lambda_1 - \{(0, 0, 0)\}$. Since f_a is homogeneous of degree 3, it follows that

$$\begin{aligned}\Delta(K_a) &= \inf\{d(\Lambda) : \inf_{\Lambda'}|f_a| \geqslant 1\}\\ &= \inf_{d(\Lambda_1)=1}\ \inf_{\Lambda_1'}\{d \in \mathbb{R} : \inf_{\Lambda_1'}|f_a| \geqslant 1/d\}\\ &= (\sup_{d(\Lambda_1)=1}\ \inf_{\Lambda_1'}|f_a|)^{-1},\end{aligned}$$

hence $M_a = \Delta(K_a)^{-1}$. We further note that the ellipsoid $E_t(r)$ can be transformed into the unit sphere by the linear transformation

$$x' = (x + ty + tz)r^{-1},\ y' = (\sqrt{1 - t^2}\ y + \sqrt{t - t^2}\ z)r^{-1},$$

$$z' = \frac{\sqrt{(1 - t)(1 + 2t)}}{r\sqrt{1 + t}}z$$

which is of determinant $(1 - t)\sqrt{1 + 2t}\ r^{-3}$. Since the critical determinant of the unit sphere equals $1/\sqrt{2}$ (see Ollerenshaw [7] or [5], p. 259), we conclude that

$$\Delta(E_t(r)) = r^3((1 - t)\sqrt{1 + 2t}\sqrt{2})^{-1}. \tag{3}$$

If we choose r maximal such that $E_t(r) \subset K_a$, then obviously $\Delta(K_a) \geqslant \Delta(E_t(r))$, hence

$$M_a = \Delta(K_a)^{-1} \leqslant r^{-3}\sqrt{2}(1 - t)\sqrt{1 + 2t}$$

and, by homogeneity,

$$\max_{E_t(r)} |f_a| = 1 \iff \max_{E_t(1)} |f_a| = r^{-3}.$$

Therefore, it suffices to establish the following

Lemma: For arbitrary t with $-\frac{1}{2} < t < 1$, $t \neq 0$, the absolute maximum of $|f_a|$ on $E_t(1)$ equals $m_a(t)$.

Proof: Since the absolute maximum of $|f_a|$ can be found among the relative extrema of f_a on the boundary of $E_t(1)$, we determine the latter by Lagrange's rule. We obtain

$$3x^2 + 3ayz + k(2x + 2t(y + z)) = 0, \tag{4}$$

$$3y^2 + 3axz + k(2y + 2t(x + z)) = 0, \tag{5}$$

$$3z^2 + 3axy + k(2z + 2t(x + y)) = 0, \tag{6}$$

$$x^2 + y^2 + z^2 + 2t(xy + xz + yz) = 1. \tag{7}$$

This system does not have any solution with $x \neq y \neq z \neq x$, for otherwise we could infer from (4) and (5) (subtracting and dividing by $x - y$) that

$$3(x + y) - 3az + 2k - 2kt = 0$$

and similarly, from (5) and (6), that

$$3(y + z) - 3ax + 2k - 2kt = 0.$$

Again subtracting, we would get the contradiction $x = z$ (at least for $a \neq -1$; the case $a = -1$ then can be settled by an obvious continuity argument).

Furthermore, it is impossible that a solution of our system satisfies $x = y = 0$, because this would imply that $ktz = 0$ and $z(3 + 2k) = 0$, hence $z = 0$, which contradicts (7). There remain two possibilities (apart from cyclic permutations).

Case 1: $x = y = z \neq 0$. By (7), we have

$$x = y = z = \pm(1 + 2t)^{-1/2}3^{-1/2}$$

and for these values of x, y, and z,

$$|f_a| = |1 + a|(1 + 2t)^{-3/2}3^{-1/2}. \tag{8}$$

Case 2: $0 \neq x = y \neq z$. Eliminating k from (4) and (6), we get

$$(2t - a - at)x^3 + (1 + at)x^2z + (a - 1 - t)xz^2 - tz^3 = 0.$$

This can be divided by $x - z$ and yields

$$tz^2 + (1 + 2t - a)xz + (2t - a - at)x^2 = 0,$$

hence $z/x = c_j$ ($j = 1, 2$, as defined in our theorem). From (7) we deduce that

$$x = y = \pm(2 + 2t + 4tc_j + c_j^2)^{-1/2}, \quad z = c_jx,$$

and for these values of x, y, and z,

$$|f_a| = \phi_j(t) \qquad (j = 1, 2). \tag{9}$$

Combining (8) and (9), we complete the proof of the lemma and thereby that of our theorem.

Concluding Remarks: Letting $t \to 0$ in our result, we just obtain Golser's theorem 1 in [4]. However, this choice of t turns out not to be the optimal one. In principle, one could look for an "advantageous" choice of the parameter t (for a given value of the constant a) by computer calculations, but it can be justified by straightforward monotonicity considerations that it is optimal to choose t such that $\max\{\phi_1(t), \phi_2(t)\}$ equals the right-hand side of (8).

We conclude the paper with tables indicating the new upper bounds for M_a (for certain values of a) as well as the corresponding "favorable" values of t and the previously-known best results due to Golser [3], [4].

a	0.9	1	2	2.9
t	0.02799	0.040786	0.07973	0.07301
$M_a \leqslant$	1.428	1.4483	1.9442	2.5758
Golser: $M_a \leqslant$	1.454	1.5018	2.0597	2.5775

a	-6	-5	-4	-3	-2	-1.2
t	-0.064204	-0.07892	-0.101987	-0.14273	-0.23042	-0.41324
$M_a \leqslant$	4.9848	4.1843	3.391	2.6116	1.8634	1.33
Golser: $M_a \leqslant$	5.03779	4.31314	3.58475	2.85169	2.1106	1.54372

REFERENCES

1. H. Davenport. "On the Minimum of a Ternary Cubic Form." *J. London Math. Soc.* *19* (1944):13-18.
2. H. Davenport. "On the Minimum of $x^3 + y^3 + z^3$." *J. London Math. Soc.* *21* (1946):82-86.
3. G. Golser. "Schranken für Gitterkonstanten einiger Sternkörper." Ph.D. Dissertation, University of Vienna, 1973.
4. G. Golser. "Über Gitterkonstanten spezieller Sternkörper." *Sitz.-Ber. Osterr. Akad. Wiss., Abt. II, math.-naturwiss. Kl.* *184* (1975):227-238.
5. C. G. Lekkerkerker. *Geometry of Numbers*. Amsterdam-London: North Holland Publishing Co., 1969.
6. L. J. Mordell. "On the Minimum of a Ternary Cubic Form." *J. London Math. Soc.* *19* (1944):6-12.
7. K. Ollerenshaw. "The Critical Lattices of a Sphere." *J. London Math. Soc.* *23* (1948):297-299.

◆◇◆◇◆