Canonical number systems for complex integers

By I. KÁTAI and J. SZABÓ in Budapest

1. It is a well-known fact that every non-negative integer N has a unique representation of the form

$$N = a_0 + a_1 A + \ldots + a_k A^k, \tag{1.1}$$

where the integers a_j are selected from the set $\{0, 1, \ldots, A-1\}$, and A is an integer, $A \geqq 2$. Furthermore, choosing a negative integer $-A$ $(A \geqq 2)$, we can represent every integer N as a sum:

$$N = a_0 + a_1(-A) + \ldots + a_k(-A)^k, \quad 0 \leqq a_j \leqq A-1 \qquad (j = 0, 1, \ldots, k-1), \tag{1.2}$$

where a_j are integers. The representation (1.2) is also unique.

The number systems of negative base have some applications in the theory of computations.

The following question seems to be interesting: Given a Gaussian integer ϑ, can we represent every Gaussian integer α in the form

$$\alpha = r_0 + r_1 \vartheta + \ldots + r_k \vartheta^k \tag{1.3}$$

or not? Here $r_j \in \mathfrak{A}$, $\mathfrak{A}$ being a fixed complete residue system mod ϑ.

If the answer is affirmative, we say that $(\vartheta, \mathfrak{A})$ is a number system.

We shall investigate only the case $\mathfrak{A} = \mathfrak{A}_0$ where

$$\mathfrak{A}_0 = \{0, 1, \ldots, N(\vartheta) - 1\}, \tag{1.4}$$

and $N(\vartheta)$ denotes the "norm"

$$N(\vartheta) = \vartheta \cdot \bar{\vartheta} = (\operatorname{Re} \vartheta)^2 + (\operatorname{Im} \vartheta)^2.$$

It is known that for $\vartheta = -1 + i$, $(\vartheta, \mathfrak{A}_0)$ is a number system; see [1]
We prove:

Theorem 1. *$(\vartheta, \mathfrak{A}_0)$ is a number system if and only if*

a) $\operatorname{Re} \vartheta < 0$ *and* b) $\operatorname{Im} \vartheta = \pm 1$.

For $\vartheta = -A \pm i$ the representation of α in the form (1.3) is unique.

Theorem 2. *Let $\vartheta=-A\pm i$, z an arbitrary complex number. Then*

$$z = a_l\vartheta^l + \dots + a_0 + \frac{a_{-1}}{\vartheta} + \frac{a_{-2}}{\vartheta^2} + \dots, \tag{1.5}$$

where $a_j \in \mathfrak{A}_0$ $(j=l, l-1, \dots, 0, -1, -2, \dots)$.

We do not assert the uniqueness of the representation of z in the form (1.5).

2. Proof of Theorem 1. *Necessity.* Let $\vartheta=A+Bi$. Then

$$\mathfrak{A}_0 = \{0, 1, \dots, A^2+B^2-1\}.$$

It is obvious that $\mathfrak{A}_0$ must be a complete residue system mod ϑ if $(\vartheta, \mathfrak{A}_0)$ is a number system. In the opposite case there is an α which is incongruent to k for every k in $\mathfrak{A}_0$, but from (1.3) $\alpha\equiv r_0 (\mathrm{mod}\ \vartheta)$, $r_0\in\mathfrak{A}_0$ follows, and this is a contradiction.

Suppose that $A>0$. We prove that $\alpha=(1-A)+iB=1-\bar{\vartheta}$ has no representation of type (1.3). Suppose in the contrary that

$$\alpha = r_0 + r_1\vartheta + \dots + r_k\vartheta^k. \tag{2.1}$$

Let

$$\varrho = \alpha(1-\vartheta) = (1-A)^2+B^2 = A^2+B^2-2A+1.$$

Since $A\geqq 1$, we have $\varrho\in\mathfrak{A}_0$. From (2.1) we get

$$\varrho = r_0 + (r_1-r_0)\vartheta + \dots + (r_k-r_{k-1})\vartheta^k - r_k\vartheta^{k+1}.$$

Hence $\varrho\equiv r_0 \bmod \vartheta$, and by $\varrho\in\mathfrak{A}_0$, $r_0\in\mathfrak{A}_0$ we get: $\varrho=r_0$. So

$$(r_1-r_0)\vartheta + \dots + (r_k-r_{k-1})\vartheta^k - r_k\vartheta^{k+1} = 0.$$

Hence it follows immediately that

$$r_1-r_0 = 0, \dots, r_k-r_{k-1} = 0, \quad r_k = 0,$$

whence $r_k=r_{k-1}=\dots=r_1=r_0=0$. Therefore $\varrho=0$, and so $A=1$, $B=0$. But it is obvious that $\vartheta=1$ is not a base of a number system. Similarly, $\vartheta=\pm i$ $(A=0, B=\pm 1)$ is not a base of a number system, either.

Let now $\mathrm{Im}\,\vartheta=B\neq\pm 1$. Let us take into account that B is a divisor of $\mathrm{Im}\,\vartheta^\nu$ $(\nu=1, 2, \dots)$. Hence, for an α of (1.3) we get:

$$\mathrm{Im}\,\alpha = r_1\,\mathrm{Im}\,\vartheta + \dots + r_k\,\mathrm{Im}\,\vartheta^k,$$

and so $B|\mathrm{Im}\,\alpha$. Consequently, (1.3) will not hold for $\alpha=i$ $(B\neq\pm 1)$.

Sufficiency. Let now $\vartheta=-A+i$ $(A\geqq 1)$. Then $\mathfrak{A}_0$ is a complete residue system mod ϑ as it is well known. Let us take into account, that

$$\vartheta^2+2A\vartheta+A^2+1 = 0. \tag{2.2}$$

Let $\alpha=E+Fi$ be an arbitrary Gaussian integer. Taking $D=F$, $C=E+AF$, we get

$$\alpha = C+D\vartheta. \tag{2.3}$$

First we prove that every α has the form

$$\alpha = U+V\vartheta+X\vartheta^2+Y\vartheta^3, \tag{2.4}$$

where U, V, X, Y are non-negative integers. From (2.2) we have

$$-1 = \vartheta^2+2A\vartheta+A^2.$$

Assuming that $C<0$ we can substitute C in (2.3) by

$$|C|\cdot\vartheta^2+2A|C|\cdot\vartheta+A^2|C|.$$

In the case $D<0$ we take a similar substitution, and get (2.4).

We shall use the following relation:

$$A^2+1 = \vartheta^3+(2A-1)\vartheta^2+(A-1)^2\vartheta. \tag{2.5}$$

Let

$$\alpha = d_0+d_1\vartheta+\dots+d_k\vartheta^k \quad (k \geqq 3), \quad d_j \geqq 0 \quad (j = 0, \dots, k). \tag{2.6}$$

Let

$$t(\alpha, d) = d_0+d_1+\dots+d_k; \tag{2.7}$$

$t(\alpha, d)$ is a non-negative integer, $t(\alpha, d)=0$ only if $\alpha=0$.

We take

$$d_0 = r_0+tN(\vartheta) = r_0+t(A^2+1),$$

$t\geqq 0$, integer, $0\leqq r_0\leqq A^2$. From (2.5) we have

$$d_0 = r_0+t(A^2+1) = r_0+t(A-1)^2\vartheta+t(2A-1)\vartheta^2+t\vartheta^3. \tag{2.8}$$

We take the right hand side of (2.8) into (2.6). Then

$$\begin{aligned}\alpha &= r_0+\big(d_1+t(A-1)^2\big)\vartheta+\big(d_2+t(2A-1)\big)\vartheta^2+(d_3+t)\vartheta^3+d_4\vartheta^4+\dots+d_k\vartheta^k = \\ &= d_0^*+d_1^*\vartheta+\dots+d_k^*\vartheta^k.\end{aligned} \tag{2.9}$$

Since

$$-t(A+1)^2+t(A-1)^2+t(2A-1)+t = 0,$$

therefore

$$t(\alpha, d^*) = d_0^*+\dots+d_k^* = t(\alpha, d), \quad d_j^* \geqq 0 \qquad (j = 0, \dots, k).$$

Let

$$\alpha_1 = d_1^*+d_2^*\vartheta+\dots+d_k^*\vartheta^{k-1}. \tag{2.10}$$

We have

$$(2.11)\qquad \alpha = \alpha_1\vartheta + r_0 \quad (r_0 \in \mathfrak{A}_0),$$

$$t(\alpha_1, d^*) = d_1^* + d_2^* + \ldots + d_k^*.$$

It is obvious that $t(\alpha_1, d^*) < t(\alpha, d)$, when $r_0 \neq 0$. For $r_0 = 0$, $t(\alpha_1, d^*) = t(\alpha, d)$.

Now we write $t(\alpha, d) = t(\alpha)$, $t(\alpha_1, d^*) = t(\alpha_1), \ldots$. We repeat the algorithm (2.9), (2.11):

$$\alpha = \alpha_1\vartheta + r_0, \quad \alpha_1 = \alpha_2\vartheta + r_1, \quad \ldots, \quad \alpha_{j-1} = \alpha_j\vartheta + r_{j-1} \quad (r_i \in \mathfrak{A}_0).$$

Then $t(\alpha) \geqq t(\alpha_1) \geqq \ldots$ and $t(\alpha_i) > t(\alpha_{i+1})$ when $r_i \neq 0$. This process is terminated at the jth step if $\alpha_j = 0$. In this case we get

$$\alpha = r_0 + r_1\vartheta + \ldots + r_{j-1}\vartheta^{j-1} \quad (r_i \in \mathfrak{A}_0).$$

Suppose that the process is not terminated. Then for a suitably large i

$$t(\alpha_i) = t(\alpha_{i+1}) = \ldots (\neq 0).$$

Hence

$$\alpha_i = \alpha_{i+1}\vartheta, \ldots \alpha_{i+k-1} = \alpha_{i+k}\vartheta$$

and, therefore, $\vartheta^k | \alpha_i$ $(k = 1, 2, \ldots)$. This holds only if $\alpha_i = 0$.

We proved the existence of the representation of α in the form (1.3).

Let us suppose now that there is an α wich has two different representations:

$$\alpha = r_0 + r_1\vartheta + \ldots + r_k\vartheta^k = s_0 + s_1\vartheta + \ldots + s_k\vartheta^k, \quad r_i, s_i \in \mathfrak{A}_0.$$

Then $0 = (r_0 - s_0) + (r_1 - s_1)\vartheta + \ldots + (r_k - s_k)\vartheta^k$ and therefore $r_0 \equiv s_0 \bmod \vartheta$; as r_0, $s_0 \in \mathfrak{A}_0$ we get $r_0 = s_0$. Dividing by ϑ, we get

$$0 = (r_1 - s_1) + \ldots + (r_k - s_k)\vartheta^{k-1}.$$

We repeat the argument. Finally we get:

$$r_0 = s_0, r_1 = s_1, \ldots, r_k = s_k.$$

We have proved the theorem for $\vartheta = -A + i$.

Let now $\vartheta = -A - i$. Using the theorem for $\bar{\vartheta} = -A + i$, we get

$$\bar{\alpha} = r_0 + r_1\bar{\vartheta} + \ldots + r_k\bar{\vartheta}^k \quad (r_i \in \mathfrak{A}_0)$$

for every Gaussian integer $\bar{\alpha}$. Hence

$$\alpha = r_0 + r_1\vartheta + \ldots + r_k\vartheta^k,$$

and so the theorem holds for $\vartheta = -A - i$, too.

3. Proof of Theorem 2. Let z be an arbitrary complex number, $z=x+iy$. Let

(3.1) $$\vartheta^k = U_k + iV_k.$$

We have

(3.2) $$z = \frac{z\vartheta^k}{\vartheta^k} = \frac{(x+iy)(U_k+iV_k)}{\vartheta^k} = \frac{C_k+D_k i}{\vartheta^k} + \frac{u_k+v_k i}{\vartheta^k},$$

where C_k, D_k are rational integers, $|u_k|<1$, $|v_k|<1$. Let

(3.3) $$z_k = \frac{C_k+iD_k}{\vartheta^k}, \quad \delta_k = \frac{u_k+iv_k}{\vartheta^k}.$$

It is obvious that $\delta_k \to 0$ $(k\to\infty)$, and so $z_k \to z$. Since C_k+iD_k is a Gaussian integer, by Theorem 1 we have

(3.4) $$C_k+iD_k = a_t^*\vartheta^t + \dots + a_0^*, \quad t = t(k).$$

First we prove that the sequence $t(k)-k$ $(k=1, 2, \dots)$ has an upper bound. Indeed, from (3.4)

$$z_k = a_t^*\vartheta^{t-k} + \dots + a_0^*\vartheta^{-k}.$$

Hence

(3.5) $$a_t^*\vartheta^{t-k} + \dots + a_k^* = z_k - \frac{a_{k-1}^*}{\vartheta} - \dots - \frac{a_0^*}{\vartheta^k},$$

and so

(3.6) $$|a_t^*\vartheta^{t-k} + \dots + a_k^*| \leqq |z_k| + \frac{a_{k-1}^*}{|\vartheta|} + \dots + \frac{a_0^*}{|\vartheta|^k} \leqq$$
$$|z| + |\delta_k| + A^2\left(\frac{1}{|\vartheta|} + \frac{1}{|\vartheta|^2} + \dots\right) \leqq |z| + |\delta_k| + \frac{A^2}{|\vartheta|-1}.$$

Hence it follows that

(3.7) $$|a_t^*\vartheta^{t-k} + \dots + a_k^*| \leqq c,$$

$c=c(z)$ being a suitable positive constant.

Since the representation of Gaussian integers in the form (1.3) is unique, and the circle $|w|\leqq c$ contains only a finite set of Gaussian integers, therefore $t(k)-k$ has an upper bound. Let K be an integer, $t-k\leqq K$. Then we can write z_k as

(3.8) $$z_k = a_K^{(k)}\vartheta^K + \dots + a_0^{(k)} + \frac{a_{-1}^{(k)}}{\vartheta} + \frac{a_{-2}^{(k)}}{\vartheta^2} + \dots,$$

where $a_j^{(k)} \in \mathfrak{A}_0$ $(j=K, K-1, \dots, 0, -1, \dots)$. Let $b_K \in \mathfrak{A}_0$ be an integer so that $a_K^{(k)} = b_K$ for infinitely many k. Let S_K be the subset of those integers k satisfying $a_K^{(k)} =$

$=b_k$. Suppose that $S_K, \ldots, S_{l+1}$ is constructed $(S_k \supseteq \ldots \supseteq S_{l+1})$. Then there is a $b_l \in \mathfrak{A}_0$, such that for infinitely many k in S_{l+1} $a_l^{(k)} = b_l$. Let S_l be the set of these k's. S_l has infinitely many elements. We repeat this argument for $K, K-1, \ldots 0, -1, \ldots$. Let

$$w = b_K \vartheta^K + \ldots + b_0 + \frac{b_{-1}}{\vartheta} + \ldots.$$

Let $k_1 < k_2 < \ldots$ be an infinite sequence, so that

$$k_\nu \in S_{K-\nu+1} \quad (\nu = 1, 2, \ldots).$$

Since

$$z_k = b_K \vartheta^K + \ldots + b_{K-\nu+1} \vartheta^{K-\nu+1} + a_{K-\nu}^{(k_\nu)} \vartheta^{K-\nu} + \ldots,$$

therefore

$$\lim_{\nu \to \infty} z_{k_\nu} = w.$$

Taking into account that $\lim\limits_{k \to \infty} z_k = z$, we have $w = z$. Hence it follows that (3.9) is a suitable representation of z.

We have proved Theorem 2.

Reference

[1] D. E. Knuth, *The art of computer programming*. Vol. 2, Addison—Wesley Publishing Company (London, 1971).

(Received December 28, 1974)