FLORENTIN SMARANDACHE

Another Integer Algorithm To Solve Linear Equations (Using Congruences)

In Florentin Smarandache: “Collected Papers”, vol. I (second edition). Ann Arbor (USA): InfoLearnQuest, 2007.

ANOTHER INTEGER ALGORITHM TO SOLVE LINEAR EQUATIONS (USING CONGRUENCES)

In this section is presented a new integer number algorithm for linear equation. This algorithm is more "rapid" than W. Sierpinski's presented in [1] in the sense that it reaches the general solution after a smaller number of iterations. Its correctness will be thoroughly demonstrated.

Another Integer Algorithm.

Let's us consider the equation (1); (the case $a_i, b \in \mathbb{Q},\ i = \overline{1,n}$ is reduced to the case (1) by reducing to the same denominator and eliminating the denominators). Let $d = (a_1, ..., a_n)$. If $d \mid b$ then the equation does not have integer solutions, while if $d \nmid b$ the equation has integer solutions (according to a well-known theorem from the number theory).

If the equation has solutions and $d \neq$ we divide the equation by d. Then $d = 1$ (we do not make any restriction if we consider the maximal co-divisor positive).

Also,

(a) If all a_i the equation is trivial; it has the general integer solution $x_i = k_i \in \mathbb{Z},\ i = \overline{1,n}$, when $b = 0$ (the only case when the general solution is n-times undetermined) and does not have solution when $b \neq 0$.

(b) If $\exists i,\ 1 \le i \le n$ such that $a_i = \pm 1$ then the general integer solution is:

$$x_i = -a_i \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} a_j k_j - b \right) \text{ and } x_s = k_s \in \mathbb{Z},\ s \in \{1, ..., n\} \setminus \{i\}$$

The proof of this assertion was given in [4]. All these cases are trivial, therefore we will leave them aside. The following algorithm can be written:

Input

A linear equation:

$$(2) \qquad \sum_{i=1}^{n} a_i x_i = b,\ a_i, b \in \mathbb{Z},\ a_i \neq \pm 1,\ i = \overline{1,n},$$

with not all $a_i = 0$ and $(a_1, ..., a_n) = 1$.

Output

The integer general solution of the equation.

Method

1. $h := 1,\ p := 1$

2. Calculate $\min\limits_{1 \le i, j \le n} \left\{ |r|,\ r \equiv a_i (\operatorname{mod} a_j),\ |r| < |a_j| \right\}$ and determine r and the pair (i, j) for which this minimum can be obtained (when there are more possibilities we have to choose one of them).

3. If $|r| \neq$ go to step 4.
If $|r| = 1$, then

$$\begin{cases} x_i := r\left(-a_j t_h - \sum\limits_{\substack{s=1 \\ s\notin\{i,j\}}}^{n} a_s x_s + b \right) \\ x_j := r\left(a_i t_h + \dfrac{a_i - r}{a_j} \cdot \sum\limits_{\substack{s=1 \\ s\notin\{i,j\}}}^{n} a_s x_s + \dfrac{r - a_i}{a_j} b \right) \end{cases}$$

(A) Substitute the values thus determined of these unknowns in all the statements (p), $p = 1, 2, ...$ (if possible).

(B) From the last relation (p) obtained in the algorithm substitute in all relations: $(\overline{p} - 1), (\overline{p} - 2), ..., (1)$

(C) Every statement, starting in order from $(\overline{p} - 1)$ should be applied the same procedure as in (B): then $(\overline{p} - 2), ..., (3)$ respectively.

(D) Write the values of the unknowns x_i, $i = \overline{1, n}$, from the initial equation (writing the corresponding integer number parameters from the right term of these unknowns with $k_1, ..., k_{n-1}$), STOP.

4. Write the statement (p): $x_j = t_h - \dfrac{a_i - r}{a_j} x_i$

5. Assign $\quad x_j := t_h \qquad h := h + 1$

$\quad a_i := r \qquad p := p + 1$

The other coefficients and variables remain unchanged go back to step 2.

The Correctness of the Algorithm

Let us consider linear equation (2). Under these conditions, the following properties exist:

Lemma 1. The set $M = \left\{ |r|,\ r \equiv a_i (\mathrm{mod}\, a_j),\ 0 < |r| < \left|a_j\right| \right\}$ has a minimum.

Proof:

Obviously $M \subset \mathbb{N}^*$ and M is finite because the equation has a finite number of coefficients: n, and considering all the possible combinations of these, by twos, there is the maximum AR_n^2 (arranged with repetition) = n^2 elements.

Let us show, by *reductio ad absurdum*, that $M \neq \emptyset$.

$M \neq \emptyset \iff a_i \equiv 0 (\mathrm{mod}\, a_j)\ \forall i, j = \overline{1, n}$. Hence $a_j \equiv 0 (\mathrm{mod}\, a_i)\ \forall i, j = \overline{1, n}$. Or this is possible only when $|a_i| = |a_j|,\ \forall i, j = \overline{1, n}$, which is equivalent to

$(a_1,..,a_n) = a_i,\ \forall i \in \overline{1,n}$. But $(a_1,..,a_n) = 1$ are a restriction from the assumption. It follows that $|a_i| = \overline{1,n},\ \forall i \in \overline{1,n}$ a fact which contradicts the other restrictions of the assumption.

$M \neq 0$ and finite, it follows that M has a minimum.

Lemma 2. If $|r| = \min\limits_{1 \le i,j \le n} M$, then $|r| < |a_i|,\ \forall i \in \overline{1,n}$.

Proof:

We assume conversely, that $\exists i_0,\ 1 \le i_0 \le n$ such that $|r| \ge \left|a_{i_0}\right|$.

Then $|r| \ge \min\limits_{1 \le j \le n}\left\{\ |a_j|\right\} = \left|a_{j_0}\right| \neq 1,\ \ 1 \le j_0 \le n$. Let $a_{p_0},\ 1 \le p_0 \le n$, such that $\left|a_{p_0}\right| > \left|a_{j_0}\right|$ and a_{p_0} is not divided by a_j^0.

There is a coefficient in the equation, $\left|a_{j_0}\right|$ which is the minimum and the coefficients are not equal among themselves (conversely, it would mean that $(a_1,..,a_n) = a_1 = \pm 1$ which is against the hypothesis and, again, of the coefficients whose absolute value is greater that $\left|a_{ij_0}\right|$ not all can be divided by a_{j_0} (conversely, it would similarly result in $(a_1,..,a_n) = a_{j_0} \neq \pm 1$.

We write $\left[a_{p_0} / a_{j_0}\right] = q_0 \in \mathbb{Z}$ (integer portion), and $r = a_{p_0} - q_0 a_{j_0} \in \mathbb{Z}$. We have $a_{p_0} \equiv r_0 (\operatorname{mod} a_{j_0})$ and $0 < |r_0| < \left|a_{j_0}\right| < \left|a_{i_0}\right| \le |r|$. Thus, we have found an r_0 which $|r_0| < |r|$ contradicts the definition of minimum given to $|r|$.

Thus $|r| < |a_i|,\ \forall i \in \overline{1,n}$.

Lemma 3. If $|r| = \min M = 1$ for the pair of indices (i, j), then:

$$\left\{\begin{array}{l} x_i = r\left(-a_j t_h - \sum\limits_{\substack{s=1 \\ s \notin \{i,j\}}}^{n} a_s k_s + b\right) \\ x_j = r\left(a_i t_h + \dfrac{a_i - r}{a_j} \cdot \sum\limits_{\substack{s=1 \\ s \notin \{i,j\}}}^{n} a_s k_s + \dfrac{r - a_i}{a_j} b\right) \\ x_s = k_s \in \mathbb{Z},\ s \in \{1,...,n\} \setminus \{i,j\} \end{array}\right.$$

is the general integer solution of equation (2).

Proof:

Let $x_e = x_e^0$, $e = \overline{1,n}$, be a particular integer solution of equation (2). Then $\exists k_s = x_s^0 \in \mathbb{Z}$, $s \in \{1,...,n\} \setminus \{i,j\}$ and $t_h = x_j^0 + \frac{a_i - r}{a_j} x_i^0 \in \mathbb{Z}$ (because $a_i - r = Ma_j$) such that:

$$x_i = r - a_j \left(x_j^0 + \frac{a_i - r}{a_j} x_i \right) - \sum_{\substack{s=1 \\ s \notin \{i,j\}}}^{n} a_s x_s^0 + b = x_i^0$$

$$x_j = r - a_j \left(x_j^0 + \frac{a_i - r}{a_j} x_i^0 \right) + \frac{a_i - r}{a_j} - \sum_{\substack{s=1 \\ s \notin \{i,j\}}}^{n} a_s x_s^0 + \frac{r - a_i}{a_j} b = x_i^0$$

and $x_s = k_s = x_s^0$, $s \in \{1,...,n\} \setminus \{i,j\}$.

Lemma 4. Let $|r| \neq$ and (i,j) be the pair of indices for which this minimum can be obtained. Again, let's consider the system of linear equations:

$$(3) \qquad \begin{cases} a_j t_h + r x_i + \sum\limits_{\substack{s=1 \\ s \notin \{i,j\}}}^{n} a_s x_s = b \\ t_h = x_j + \dfrac{a_i - r}{a_j} x_i \end{cases}$$

Then $x_e = x_e^0$, $e = \overline{1,n}$ is a particular integer solution for (2) if and only if $x_e = x_e^0$, $e \in \{1,...,n\} \setminus \{j\}$ and $t_h = t_h^0 = x_j^0 + \frac{a_i - r}{a_j} x_i$ is the particular integer solution of (3).

Proof:

$x_e = x_e^0$, $e = \overline{1,n}$ is a particular solution for (2) if and only if

$$\sum_{e=1}^{n} a_e x_e^0 = b \Leftrightarrow \sum_{\substack{s=1 \\ s \notin \{i,j\}}}^{n} a_s x_s^0 + a_j \left(x_j^0 + \frac{a_i - r}{a_j} x_i^0 \right) + r x_i^0 = b \Leftrightarrow$$

$$\Leftrightarrow a_j t_h^0 + r x_i^0 + \sum_{\substack{s=1 \\ s \notin \{i,j\}}}^{n} a_s x_s^0 = b \quad \text{and} \quad t_h^0 = x_j^0 + \frac{a_i - r}{a_j} x_i^0 \in \mathbb{Z} \quad \Leftrightarrow x_e = x_e^0,$$

$e \in \{1,...,n\} \setminus \{j\}$ and $t_h = t_h^0$ is a particular integer solution for (3).

Lemma 5. The previous algorithm is finite.

Proof:

When $|r| = 1$ the algorithm stops at step 3. We will discuss the case when $|r| \neq 1$. According to the definition of r, $|r| \in \mathbb{N}^*$. We will show that the row of $r - s$ successively obtained by following the algorithm several times is decreasing with cycle, and each cycle is not equal to the previous, by 1. Let r_1 be the first obtained by following the algorithm one time. $|r_1| \neq 1$ then go to step 4, and then step 5. According to lemma 2, $|r_1| < |a_i|$, $\forall i = \overline{1,n}$.

Now we shall follow the algorithm a second time, but this time for an equation in which r_1 (according to step 5) is substituted by a_i. Again, according to lemma 2, the new $|r|$ written $|r_2|$ will have the propriety: $|r_2| < |r_1|$. We will get to $|r| = 1$ because $|r| \geq 1$ and $|r| < \infty$, and if $|r| \neq 1$, following the algorithm once again we get $|r| < |r_1|$ and so on. Hence, the algorithm has a finite number of repetitions.

Theorem of Correctness. The previous algorithm calculates the general solution of the linear equation correctly (2).

Proof:

According to lemma 5 the algorithm is finite. From lemma 1 it follows that the set M has a minimum, hence step 2 of the algorithm has meaning. When $|r| = 1$ it was shown in lemma 3 that step 3 of the algorithm calculates the general integer solution of the respective equation correctly the equation that appears at step 3). In lemma 4 it is shown that if $|r| \neq 1$the substitutions steps 4 and 5 introduced in the initial equation, the general integer solution remains unchanged. That is, we pass from the initial equation to a linear system having the same general solution as the initial equation. The variable h is a counter of the newly introduced variables, which are used to successively decompose the system in systems of two linear equations. The variable p is a counter of the substitutions of variables (the relations, at a given moment between certain variables).

When the initial equation was decomposed to $|r| = 1$, we had to proceed in the reverse way, i.e. to compose its general integer solution. This reverse way is directed by the sub-steps 3(A), 3(B) and 3(C). The sub-step 3(D) has only an aesthetic role, i.e., to have the general solution under the form: $x_i = f_i(k_1, ..., k_{n-1})$, $i = \overline{1,n}$, f_i being linear functions with integer number of coefficients. This "if possible" shows that substitutions are not always possible. But when they are we must make all possible substitutions.

Note 1. The previous algorithm can be easily introduced into a computer program.

Note 2. The previous algorithm is more "rapid" than that of W. Sierpinski's [1], i.e., the general integer solution is reached after a smaller number of iterations (or, at least, the same) for any linear equation (2).

In the first place, both methods aim at obtaining the coefficient ± 1 for at least one unknown variable. While Sierpinski started only by chance, decomposing the greatest coefficient in the module (writing it under the form of a sum between a multiple of the following smaller coefficient (in the module) and the rest), in our algorithm this decomposition is not accidental but always seeks the smallest $|r|$

and also choose the coefficients a_i and a_j for which this minimum is achieved. That is, we test from the beginning the shortest way to the general integer solution. Sierpinski does not attempt to find the shortest way; he knows that his method will take him to the general integer solution of the equation and is not interested in how long it will take. However, when an algorithm is introduced into a computer program it is preferable that the process time should be as short as possible.

Example 1.

Let us solve in $\mathbb{Z}^3$ the equation $17x - 7y + 10z = -12$.

We apply the former algorithm.

1. $h = 1, p = 1$

2. $r = 3,\ i = 3,\ j = 2$

3. $|3| \neq 1$ go on to step 4.

4. (1) $y = t_1 - \dfrac{10-3}{-7} \cdot z = t_1 + z$

5. Assign

$$y := t_1 \qquad h := 2$$
$$a_3 := 3 \qquad p := 2$$

with the other coefficients and variables remaining unchanged, go back to step 2.

2. $r = -1,\ i = 1,\ j = 3$

3. $|-1| = 1$

$$x = -1(-3t_2 - (-7t_1) - 12) = 3t_2 - 7t_1 - 12$$
$$z = -1\left(17t_2 + (-7t_1) \cdot \frac{17-(-1)}{3} + \frac{-1-17}{3}(-12)\right) = 17t_2 + 42t_1 - 72$$

(A) We substitute the values of x and z thus determined into the only statement (p) we have:

(1) $y = t_1 + z = --17t_2 + 43t_1 - 72$

(B) The substitution is not possible.

(C) The substitution is not possible.

(D) The general integer solution of the equation is:

$$\begin{cases} x = 3k_1 - 7k_2 + 12 \\ y = -17k_1 + 43k_2 - 72 \\ z = -17k_1 + 42k_2 - 72; \qquad k_1, k_2 \in \mathbb{Z} \end{cases}$$

REFERENCES:

[1] Sierpinski, W, - Ce ştim şi ce nu ştim despre numerele prime? - Editura Stiinţifică, Bucharest, 1966.

[2] Creangă, I., Cazacu, C., Mihuţ, P., Opaiţ, Gh., Corina Reisher – Introducere în teoria numerelor, Ed. Did. şi Ped., Bucharest, 1965.

[3] Popovici, C. P. – Aritmetica şi teoria numerelor, Ed. Did. şi Ped., Bucharest, 1963.

[4] Smarandache, Florentin – Un algoritm de rezolvare în numere întregi a ecuaţiilor liniare.