DIMACS Technical Report 2005-29
January 2006

Universal Subsets of Z_n, Linear Integer Optimization, and Integer Factorization

by

Zhivko Nedev[1]

[1]Univ. of Victoria, `znedev@math.uvic.ca`. Part of the work was done while visiting DIMACS.

ABSTRACT

We consider two classes of sets in Z_n. A non-empty subset U of Z_n is universal (the first class) if for all $x \in U$, and for all $0 < l \leq n/2$ at least one of $x \pm l \pmod{n}$ lies in U. For each universal U its complement, $Z_n \backslash U$, is from the second class and vice versa. We define $\beta(n)$ to be the minimum cardinality of an universal set modulo n. Completely characterizing all sets in the second class we derive a formula for $\beta(n)$.

We demonstrate that universal sets arise in the context of a two-player game that was analyzed for the first time in [3] and has interesting connections to the prime factorization of n. Finally we model our optimization problem, *find* $\beta(n)$, as an integer linear program.

1 Introduction

In this paper we prove equivalence between a well known problem – integer factorization – and two new optimization problems. One of them has as a domain subsets of Z_n with special properties. The other one is a linear integer optimization problem. We also show a connection between these problems and a two-player game for the first time analyzed in [3].

For a standard exposition of the properties of Z_n and the integer factorization problem see [5]. For classical algorithms for integer factorization see [1] and [2]. The success of RSA cryptographic algorithm [6] is based on the assumption that it is hard to factor an arbitrary integer. For a more recent approach towards factoring large integers see [8]. A standard text on Linear Integer Optimization is [7].

The input parameter to all problems considered in this paper is a positive integer $n > 1$. As usual we denote by Z_n the set $\{0, 1, 2, \ldots, n-1\}$ with its two operations: addition and subtraction modulo n. We will be interested in two special classes of subsets in Z_n.
Definition. A nonempty subset of $\{0, \ldots, n-1\}$ will be called *universal modulo n* if for each element x and for each (integer) magnitude l, with $0 < l \leq n/2$, there is a direction $d \in \{-1, +1\}$ such that $x + d \times l \pmod{n}$ is also from the set. □

Trivial examples of universal subsets of Z_n are: Z_n for every n, and $Z_n \backslash \{i\}$ for $\forall i \in Z_n$ when n is odd. Notice that any universal set has at least $1 + \lfloor \frac{n}{2} \rfloor$ elements.
Problem 1. *For an integer $n > 1$, what is the size of a smallest universal set modulo n? How can such a set be constructed?*

We define $\beta(n)$ to be the minimal size of a universal set modulo n.
Definition. A proper subset of $\{0, \ldots, n-1\}$ will be called *middle-inclusive modulo n* if it is closed under taking midpoints. That is $M \subset Z_n$ is middle-inclusive if for all (not necessarily distinct) $a, b \in M$, each solution to the equation $2x = a + b \pmod{n}$ is also from M. □

Notice that when $a = b$, the above equation $2x = a + b \pmod{n}$ has a non-trivial solution (exactly one) only when n is even. In this case the non-trivial solution is $a + n/2$. When a, b are distinct the equation $2x = a + b \pmod{n}$ has zero, one or two solutions: if n is odd exactly one solution; when n is even, either zero, or two solutions.

Trivial examples of middle-inclusive sets are: $\varnothing$, and $\{i\}$ for $\forall i$, if n is odd.
Lemma 1.1. *For any integer $n > 1$, S is universal set modulo n if and only if $Z_n \backslash S$ is middle-inclusive set modulo n.*

Proof. ($\Rightarrow$) Let S be any universal subset of Z_n. Let $\overline{S} = Z_n \backslash S$. Obviously $| \overline{S} | < n$. Suppose $a, b \in \overline{S}$, $x \notin \overline{S}$ and $2x = a + b \bmod n$. Take the smaller of $x - a \pmod{n}$ and $-x + a \pmod{n}$, say $x - a \pmod{n} \leq n/2$. From the definition of universal set (it follows that) at least one of $x - (x - a) \pmod{n} = a$ or $x + (x - a) = 2x - a \pmod{n} = b$ is in S, which is a contradiction. Therefore $\overline{S}$ is a middle-inclusive subset of Z_n.

($\Leftarrow$) Now let S be any middle-inclusive subset of Z_n. Let $\overline{S} = Z_n \backslash S$. Obviously $| \overline{S} | \geq 1$. Let x and l be such that $x \in \overline{S}$, $0 < l \leq n/2$, and both $x \pm l \pmod{n} \in S$. Then

$2x = (x+l)+(x-l) \pmod{n}$. From the definition of middle-inclusive subset it follows that x should be in S, which is a contradiction. Therefore $\overline{S}$ is universal. □

Corollary 1. *For any $n > 1$, $\beta(n)$, the minimal size of an universal set modulo n, equals n minus the maximal size of a middle-inclusive set modulo n,* i.e. *$\beta(n) = n - \max|M|$, where M runs over all middle-inclusive sets modulo n.*

2 Formula for $\beta(n)$

Theorem 2.1. *For all $n \in Z^+$*

$$\beta(n) = \begin{cases} n & \text{if } n = 2^k \text{ for some } k. \\ \\ \frac{p-1}{p} \cdot n & \text{where } p \text{ is the smallest odd prime factor of } n. \end{cases}$$

We will prove the above theorem by completely characterizing for any $n \in Z^+$ all middle-inclusive subsets of Z_n with at least one element.

Definition. Let d and r be integers, with $d \mid n$, $d > 0$, and $0 \le r < d$. We will denote by $C_n(r,d)$ the subset of Z_n that is the arithmetic progression starting at r and having a common difference d. That is $C_n(r,d) = \left\{r + i \cdot d \mid 0 \le i < \frac{n}{d}\right\}$. □

It is convenient to have the following description. Suppose we have a round table with n positions labeled as $0, 1, \ldots, n-1$ in a clockwise manner. If $i, j \in Z_n$ are two such positions, then we denote with $d_+(i,j)$ the number of positions that have to be passed if we travel around the table clockwise from position i to position j. Obviously if $i < j$ then $d_+(i,j) = j - i$, and $d_+(j,i) = n - j + i$.

The following lemma characterizes all middle-inclusive subsets of Z_n.

Lemma 2.2. *Let $n > 1$. If $d > 1$ is an odd divisor of n, then $C_n(r,d)$ is a middle-inclusive subset of Z_n for any r. Conversely if M is a non-empty middle-inclusive subset of Z_n, then there are integers r and d with: d odd, $d > 1$, $d \mid n$, and $0 \le r < d$ such that $M = C_n(r,d)$.*

Proof. ($\Rightarrow$) Let d be odd, $d > 1$, and $d \mid n$. Let also r satisfy $0 \le r < d$. Let $S = C_n(r,d)$. Since $d > 1$ we have that $\mid S \mid < n$.

Let a and b be (not necessarily distinct) from S and let x be any solution to the equation $2x = a + b \pmod{n}$. Then $a = r + i \times d$, $b = r + j \times d$, and $2x = 2r + (i+j)d \pmod{n}$. Therefore x is either $r + \frac{i+j}{2} \times d \pmod{n}$ or $r + \frac{n+i+j}{2} \times d \pmod{n}$. In either case $x \in S$ and it follows that S is middle-inclusive.

($\Leftarrow$) Let M be any non-empty middle-inclusive subset of Z_n. We consider two cases: $\mid M \mid = 1$ and $\mid M \mid \ge 2$.

Case 1: $\mid M \mid = 1$. Let $M = \{r\}$. Then n must be odd, otherwise M would have at least two elements: r and $r + n/2$. So the conditions of the lemma hold.

Case 2: $\mid M \mid \geq 2$. Let the elements of M be sorted in increasing order: $M = \{0 \leq i_1 < i_2 < i_3 < \ldots < i_l \leq n-1\}$. By definition $\mid M \mid < n$, so $l < n$. Take any three "consecutive" elements of M, for example: i_{j-1}, i_j, i_{j+1}. Here the index arithmetic is done modulo l, which for example means that i_{l-1}, i_l, i_1 are three consecutive elements of M.

If $d_+(i_{j-1}, i_j)$ is even then $i_{j-1} + \frac{d_+(i_{j-1}, i_j)}{2} \pmod n \in Z_n$, but it is not in M. This is a contradiction because M is middle-inclusive. It follows that $d_+(i_{j-1}, i_j)$ must be odd for all j.

But then $d_+(i_{j-1}, i_{j+1}) = d_+(i_{j-1}, i_j) + d_+(i_j, i_{j+1})$ must be even. If $d_+(i_{j-1}, i_j) \neq d_+(i_j, i_{j+1})$, then the middle point from i_{j-1} to i_{j+1} is from Z_n and is different from i_j. It should be then in M but it is not, a contradiction. Therefore $d_+(i_{j-1}, i_j) = d_+(i_j, i_{j+1})$ for all j. It follows that $d_+(i_{j-1}, i_j) = d_+(i_{k-1}, i_k)$ all j, k. Let d be the common value (which is odd as noted above) for $d_+(i_{j-1}, i_j)$.

It follows that $M = \{i_1, i_2, \ldots, i_l\} = \{i_1, i_1 + d, i_1 + 2 \cdot d, \ldots, i_1 + (\frac{n}{d} - 1) \cdot d\}$. □

Proof of Theorem 2.1

Case 1: $n = 2^k$ for some $k > 0$. Suppose there exists a non-empty middle-inclusive subset M of Z_n. From the characterization lemma there exists an odd d, with $d > 1$ and $d \mid n$, which is a contradiction. Therefore the only possible middle-inlcusive subset of Z_n is the empty one. It follows then that the only possible universal subset is $S = Z_n$. Thus $\beta(n = 2^k) = n$.

Case 2: $n = d_1 \cdot 2^f$, where $d_1 \mid n$ is odd and $d_1 \geq 3$.

Since from the characterization lemma all non-empty middle-inclusive subsets M are of the form $C_n(r, d)$, where $d \mid n$ is odd and $d > 1$ we want to find the largest such subset. Obviously $C_n(r, p)$ where p is the smallest odd prime factor of n has the biggest cardinality. Notice that $\mid C_n(r, p) \mid = \frac{n}{p}$ irrespective of r, $0 \leq r < p$. Therefore $\beta(n) = n - |C_n(r, p)|$ or $\beta(n) = n - \frac{n}{p} = \frac{p-1}{p} \cdot n$, when $n \neq 2^k$. □

Observation

Obviously, if we agree that finding the smallest odd prime factor of an integer and integer factorization are equivalent problems, then we want to compute $\beta(n)$. This is because knowing p, the smallest prime odd factor of n, we can find $\beta(n) = \frac{p-1}{p} \cdot n$ and vice versa.

3 Connection with the Nagger-Mover game

Here we establish a connection between the universal subsets of Z_n and the following two-player game that was first analyzed in [3] (where it was called The Nagger-Mover game). The game is played at a circular table with n seats consecutively labelled 0 to $n-1$. The two players are called the Nagger and the Mover. If the current position is i, a round consists of the Nagger calling a magnitude ℓ with $0 < \ell \leq n/2$, after which the Mover calls a direction ($+$ or $-$). The position is then updated to $i + \ell \bmod n$ or $i - \ell \bmod n$ according to whether the Mover called $+$ or $-$. Nagger's aim in the game is to maximize the cardinality of the set of all positions occupied in the course of the game (while Mover's is to minimize it). In [3], a simple formula was given (in terms of the prime factorization of n) for the size of such a set (the function $f^*(n)$ was used for the size) if both players play optimally. Here we give a simpler proof than the one given in [3], for the formula for the function $f^*(n)$ (we will prove $f^*(n) = \beta(n)$).

We claim that $\beta(n)$ is precisely the eventual size of the occupied set if both players play optimally. To see this, first let U be any universal set that contains the current position. Consider the following strategy for the Mover. At each turn when presented with a pair (x, ℓ) consisting of the current position x and the magnitude ℓ selected by the Nagger, the Mover chooses a direction so that the next position is also from U. This is always possible, because U is universal and at least one of $x \pm \ell \pmod{n}$ is in U. Since any universal set may be translated to contain the initial position, the Mover has a strategy to ensure that no more than $\beta(n)$ positions are occupied, irrespective of Nagger's strategy.

Conversely, consider the following strategy for the Nagger. At each turn he is presented only with the current position x. The strategy for the Nagger is to choose (if possible) such an ℓ that both $x \pm \ell \pmod{n}$ have not been visited yet. If such an ℓ does not exist then he chooses ℓ sequentially to be $1, 2, \ldots, \lfloor \frac{n}{2} \rfloor, 1, 2, \ldots, \lfloor \frac{n}{2} \rfloor, \ldots$. We consider the set S of positions that arise infinitely often in the sequence of plays. Since there are finitely many positions, the set S is non-empty. We claim that S is universal. To see this, note that for any $x \in S$, the Nagger will choose any of the magnitudes $0 < \ell \leq n/2$ infinitely many times. Consequently the Mover will be presented with the pair (x, ℓ) where $x \in S$ and ℓ is a magnitude from 1 to $\lfloor \frac{n}{2} \rfloor$ infinitely many times. Therefore at least one of $x \pm \ell \pmod{n}$ will be visited infinitely many times and as such it belongs to S. We see then that S must be universal. So the set of positions visited infinitely often has cardinality at least $\beta(n)$.

4 Linear Integer Optimization problem

Now we will show equivalence between a linear integer optimization problem and our problem 1:

$$\texttt{find } \mathit{min} \mid S \mid \texttt{, where } S \texttt{ is any universal subset of } Z_n \texttt{.}$$

Let $S \subseteq Z_n$ be any optimal solution for the above optimization problem. Let x_i, for $i = 0, 1, \ldots, n-1$ be n binary variables such that:

$$x_i = \begin{cases} 1 & \text{if } i \in S \\ 0 & \text{if } i \notin S \end{cases}$$

Then $\mid S \mid = x_0 + x_1 + \ldots + x_{n-1}$ and the optimization goal, min $\mid S \mid$, in (problem 1) becomes:

$$min \;\; x_0 + x_1 + \ldots + x_{n-1}$$

Now we have to model that S is universal, i.e. for all $x \in S$, and for all integer l, with $0 < l \leq n/2$, at least one of $x \pm l \pmod n$ lies in S. It is equivalent to:

> if $x_i = 1$ then for each $j = 1 ... \lfloor \frac{n}{2} \rfloor$ at least one of the two variables x_{i+j} and x_{i-j} must be 1.

Here the addition and the subtraction operations in the indices are by modulo n arithmetic.

One way to model this constraint is by $x_i(x_{i+j} + x_{i-j}) \geq x_i$. If $x_i = 0$ this inequality is trivially satisfied: there is no constraint on x_{i+j} and x_{i-j} coming from index i. And if $x_i = 1$ then this constraint (and $\lfloor \frac{n}{2} \rfloor - 1$ more) becomes $x_{i+j} + x_{i-j} \geq 1 \Rightarrow$ at least one of x_{i+j}, x_{i-j} is 1.

Since this is a nonlinear constraint we would like to replace it by a linear one if possible (see [4]). Because x_i, x_{i+j} and x_{i-j} are boolean variables it can be achieved in several ways. The simplest possible maybe is $x_{i+j} + x_{i-j} \geq x_i$.

Here is the equivalent to Problem 1 linear integer optimization problem with an input parameter $n \in Z^+$.
Problem 2.

$$min \; x_0 + x_1 + x_2 + ... + x_{n-1} \quad \textit{subject to:}$$

$$\left.\begin{array}{r} x_{0+1} + x_{n-1} \geq x_0 \\ x_{0+2} + x_{n-2} \geq x_0 \\ ... \\ x_{0+\lfloor \frac{n}{2} \rfloor} + x_{n-\lfloor \frac{n}{2} \rfloor} \geq x_0 \end{array}\right\} \quad group \;\; of \;\; constraints \;\; for \;\; x_0$$

$$\left.\begin{array}{r} x_{1+1} + x_{1-1} \geq x_1 \\ x_{1+2} + x_{n-1} \geq x_1 \\ ... \\ x_{1+\lfloor \frac{n}{2} \rfloor} + x_{n-\lfloor \frac{n}{2} \rfloor+1} \geq x_1 \end{array}\right\} \quad group \;\; of \;\; constraints \;\; for \;\; x_1$$

$$...$$

$$\left.\begin{array}{r} x_{n-1+1} + x_{n-1-1} \geq x_{n-1} \\ x_{n-1+2} + x_{n-1-2} \geq x_{n-1} \\ ... \\ x_{n-1+\lfloor \frac{n}{2} \rfloor} + x_{n-1-\lfloor \frac{n}{2} \rfloor} \geq x_{n-1} \end{array}\right\} \quad group \;\; of \;\; constraints \;\; for \;\; x_{n-1}$$

$$x_0 + x_1 + x_2 + ... + x_{n-1} \geq 1$$

The last inequality is equivalent to the non-emptiness condition in the definition for universal subsets of Z_n: $\mid S \mid \geq 1$. There is a trivial observation: the number of variables that are

1 in any feasible solution must be at least $1 + \lfloor \frac{n}{2} \rfloor$. So the last constraint can be replaced with:

$$x_0 + x_1 + x_2 + ... + x_{n-1} \geq 1 + \lfloor \frac{n}{2} \rfloor$$

This follows directly from the last constraint: at least one variable $x_i = 1$ for some i and from the group of inequalities for that x_i.

For example for $n = 3$ the linear integer optimization problem becomes:

$$\begin{aligned} &min\ x_0 + x_1 + x_2 \quad \text{subject to:} \\ &\qquad x_1 + x_2 \geq x_0 \,\} \ group\ of\ constraints\ for\ x_0 \\ &\qquad x_2 + x_0 \geq x_1 \,\} \ group\ of\ constraints\ for\ x_1 \\ &\qquad x_0 + x_1 \geq x_2 \,\} \ group\ of\ constraints\ for\ x_2 \\ &\qquad x_0 + x_1 + x_2 \geq 1 \end{aligned}$$

There are n boolean variables and n groups of constraints. Each group has $\lfloor \frac{n}{2} \rfloor$ inequalities. One can easily see that the problem is completely symmetric for all the variables: $x_0, x_1, \ldots, x_{n-1}$.

We have proved the following:

Theorem 4.1. *The above 0,1 minimization problem has an optimal value for the function* $\sum_{i=0}^{n-1} x_i$ *as follows:*

a) if $n = 2^k$ *then the optimal value is* $min = 2^k = n$

b) if $n = 2^{\alpha_1} p^{\alpha_2} p_3^{\alpha_3} \cdots p_l^{\alpha_l}$*, where* p *is the smallest prime factor of* n *bigger than 2, then the optimal value is* $min = \frac{p-1}{p} \cdot n$

c) (corollary of b) if n *is a prime then* $min = n - 1$

Solving efficiently the above optimization problem leads to a factorization algorithm: if p is the smallest factor (of n) $\neq 2$ and min is the minimal value for the goal function then

$$p = \frac{n}{n - min}$$

Corollary 2. *Primality testing: taking only the constraints from the above optimization problem, and adding one more constraint:*

$$x_0 + x_1 + x_2 + ... + x_{n-1} \leq n - 2$$

we can ask: is there a feasible solution for this set of constraints? Obviously this is equivalent to answering if n *is prime.*

5 Open Problems

One of the reasons we can not solve efficiently Problem 2 is that the number of variables and the number of constraints is exponential in the number of bits needed to store n. Therefore we state the following open problems.

Q1: For any $n \in Z^+$ what is the linear integer optimization problem with minimal complexity such that the minimal value of its optimization function gives the smallest prime factor of n?

Q2: Is there a polynomial algorithm(of $\log_2 n$ running time) to solve the above $\{0,1\}$ linear optimization problem for certain types of n , or can we prove a lower bound of non-polynomial type?

6 Acknowledgments

We are very grateful to the following people and places for their advice, discussions and support: Gary MacGillivray, Anthony Quas, Valerie King and University of Victoria; Uriel Feige; S. Muthu Muthukrishnan and the DIMACS Center; Jeffrey O. Shallit and the School of Computer Science at University of Waterloo.

References

[1] Eric Bach and Jeffrey Shallit, *Algorithmic Number Theory*, The MIT Press, 1996.

[2] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein, *Introduction to Algorithms*, The MIT Press, 2001.

[3] Zhivko Nedev and S. Muthukrishnan, *The Nagger-Mover Game*, submitted to SIAM Journal on Discrete Mathematics, 2005.

[4] Nikola S. Nikolov, *personal communication*, University of Limerick, Ireland

[5] Ivan Niven, Herbert Zuckermann and Hugh Montgomery, *An Introduction to the Theory of Numbers*, John Wiley and Sons Inc, 1991.

[6] R. L. Rivest, A. Shamir and L. Adleman, *A Method for Obtaining Digital Signatures and Public-key Cryptosystems*, Communications of the ACM 21,2 (Feb. 1978), 120–126.

[7] Alexander Schrijver, *Theory of Linear and Integer Programming*, John Wiley and Sons Inc, 1998.

[8] A. Shamir, E. Tromer: *Factoring Large Number with the TWIRL Device*. In: Cripto 2003. Volume 2729 of LNCS, Springer (2003) 1-26.