A Probabilistic Approach on Estimating the Number of Modular Sonar Sequences

Ki-Hyeon Park and Hong-Yeop Song

Department of Electrical and Electronic Engineering
Yonsei University, Seoul, 121-749, Korea
{kh.park, hysong}@yonsei.ac.kr

Abstract. We report some results of an extensive computer search for $m \times n$ modular sonar sequences and estimate the number of inequivalent examples of size $m \times n$ using a probabilistic approach. Evidence indicates strongly that a full size example exists with extremely small probability for large m.

1 Introduction

A sonar sequence is an integer sequence that has some interesting properties for use in communication applications. Its mathematical concept was well described in [2] and the original motivation and application to some communication problems can be found in [1,6].

Recently, [8] has discussed a search for a 35×35 modular sonar sequence and, in general, $m \times m$ examples where $m = p(p+2)$ is a product of twin primes. It was for their application to the design of CDMA sequences, but they failed to find any single example beyond $m = 15$. This paper is an attempt to continue this effort, and shows some results of an extensive computer search for small values of m. Based on the search result, we now believe that no $m \times (m+1)$ modular sonar sequence exists, except for those given by the algebraic constructions. To explain this, we use some probabilistic approaches for estimating the number of $m \times (m+1)$ modular sonar sequences.

An $m \times n$ sonar sequence is defined as a function from the set of integers $\{1, 2, ..., n\} \stackrel{\triangle}{=} A_n$ to the set of integers $\{1, 2, ..., m\} \stackrel{\triangle}{=} A_m$ with the following distinct difference property (DDP) [3].

Definition 1. *(DDP) A function $f : A_n \to A_m$ has a distinct difference property if for all integers h, i, and j, with $1 \le h \le n-1$ and $1 \le i, j \le n-h$,*

$$f(i+h) - f(i) = f(j+h) - f(j) \quad \textit{implies} \quad i = j. \tag{1}$$

An $m \times n$ sonar sequence is a function $f : A_n \to A_m$ with DDP. The main problem in sonar sequences research is to determine the maximum value n for each given m such that an $m \times n$ sonar sequence exists. For values of m up to 100, the best known n is reported in [3]. To obtain these values, they have introduced "modular sonar sequences." A modular sonar sequence is a sonar

S.W. Golomb et al. (Eds.): SETA 2008, LNCS 5203, pp. 42–50, 2008.

sequence $f : A_n \to A_m$ with the condition (1) replaced by the distinct modular difference property (DMDP):

$$f(i+h) - f(i) = f(j+h) - f(j) \pmod{m} \quad \text{implies} \quad i = j. \tag{2}$$

Note that, obviously, DMDP implies DDP, but not conversely. A trivial upper bound on the maximum length n for a given m for sonar sequences is $2m$, since the maximum number of differences with $h = 1$ in (1) is $2m - 1$. Similarly for modular sonar sequences, this upper bound is given as $m+1$ since the maximum number of differences with $h = 1$ in (2) is m.

If an f is a modular sonar sequence, the function g given by

$$g(i) = uf(i) + si + a, \quad i = 1, 2, ..., n \tag{3}$$

is also a modular sonar sequence for all integer s and a, and for all integer u relatively prime to m [3]. Two $m \times n$ modular sonar sequences with this relation are said to be equivalent.

There are essentially three algebraic methods constructing an $m \times (m + 1)$ modular sonar sequence for certain values of m. These are Quadratic Method [4] and Extended Exponential Welch Method [7] both for m being a prime and Shift Sequence Method [5] for m being one less than a prime power.

Given an $m \times n$ (modular) sonar sequence, we can always have $m \times (n - 1)$ (modular) sonar sequence by deleting the last term. It works since the condition (1) or (2) remains satisfied when the domain of f is restricted to $\{1, 2, ..., n-1\}$. We call it "Reduction." Conversely, if there is no $m \times n$ (modular) sonar sequence, then there is no $m \times (n + 1)$ (modular) sonar sequence.

2 Back-Track Search and Results

This section reports some results from an exhaustive back-track search for $m \times n$ modular sonar arrays for some small values of m.

The algorithm recursively builds up a set of GOOD symbols for the current position t based on a modular difference triangle (MDT) of depth $t - 2$ constructed from the sequence $f(1), f(2), ..., f(t - 1)$ of length $t - 1$ in order to

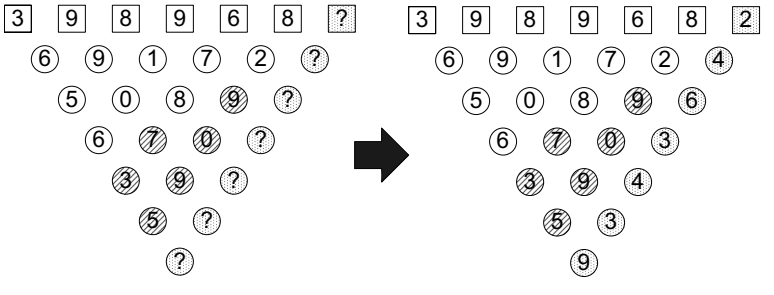

Fig. 1. Determining $f(7)$ by back-track algorithm for m=10

assign a symbol to $f(t)$. This is done by removing all the symbols from A_m, which will violate DMDP when it reaches the t-th position. When this set of GOOD symbols for the current position becomes empty, the algorithm will output the sequence constructed so far (if it has a new longer length) and then back-track. The algorithm will stop when the set of GOOD symbols for the first position becomes empty.

Figure 1 shows a situation for $m = 10$, in which the algorithm has filled up 6 terms and seeks to assign a symbol to $f(7)$. Symbols (or numbers) in squares at the top are the sequence $f(i)$ for $i = 1, 2, ..., 6$, and those in circles at level h deep are the differences mod 10 of terms in the distance h, i.e., $f(i + h) - f(i)$ (mod 10). They are said to form a modular difference triangle (MDT) with no symbol repeating in any row (except for the top row that corresponds to the sequence itself). The algorithm will remove a symbol from A_{10} (to build a set of GOOD symbols for $f(7)$) if it does not satisfy DMDP with respect to the given MDT of depth 5. Observe for example in this case that the symbol 4 will be removed since $f(2) - f(1) = 6 = 4 - f(6) \pmod{10}$ from the rows of level 1 deep. Similarly, because of the differences 9, 1, 7, and 2 at level 1 deep, the symbols 7, 9, 5 and 10 will also be removed. Because of the differences $5, 0, 8, 9$ at level 2 deep and $f(5) = 6$, the symbols $1, 6, 4, 5$ will also be removed. We note that the symbol 5 had already been removed.

Fact 1. Observe that the difference 9 at level 2 does not produce any new constraint because if $f(7) - f(5) = 9 = f(6) - f(4)$ then $f(7) - f(6) = f(5) - f(4)$. Thus, the update process will become simpler when it considers only those differences in the un-shaded circles.

We have focused on all existing examples of maximum length except for those given by the algebraic constructions mentioned in the previous section and their reductions. The initial search was to answer the following two questions:

Q.1. Determine the maximum length n_{max} such that an $m \times n_{max}$ modular sonar array exists. What would be the maximum length n_e if we count only those that are NOT equivalent to any examples constructed by the three algebraic methods mentioned in the previous section and/or their reductions?

Q.2. How many inequivalent sequences of length n_e exist for a given m, excluding those which are equivalent to the one given by the three algebraic constructions and/or their reductions?

The result of the search is shown in Table 1, from which we were able to detect the following behaviors of values m and n_e.

Observation 1. $m - n_e$ is monotonically non-decreasing as m is increasing.

Observation 2. The number of inequivalent sequences of length n_e (not equivalent to ones from the algebraic constructions) is decreasing as m is increasing for the range where the value $m - n_e$ remains the same.

Next section will be devoted to describing the above behaviors and more.

Table 1. Result of an initial search

m	Description	n_{max}	n_e	$m - n_e$	Answer to **Q.2**
5	Q,EEW	6	0	5	0
6	SS	7			
7	Q,EEW	8	8	-1	1
8	SS	9			
9	U	10	10	-1	3
10	SS	11			
11	Q,EEW	12	11	0	30
12	SS	13			
13	Q,EEW	14	13	0	17
14	U	14	14	0	2
15	SS	16	15	0	1
16	SS	17	16	0	1
17	Q,EEW	18	16	1	33
18	SS	19			
19	Q,EEW	20	17	2	321
20	U	18	18	2	136
21	U	19	19	2	17
22	SS	23			
23	Q,EEW	24			
24	SS	25			
25	U	22	22	3	4
26	SS	27			
27	U	?			
28	SS	29			
29	Q,EEW	30			
30	SS	31			
31	Q,EEW,SS	32			
32	U	?			
33	U	?			
34	U	?			
35	U	?			

SS:Shift Sequence **Q**:Quadratic **EEW**:Extended Exponential Welch **U**:Unidentified

3 Probabilistic Approach

The idea is in the back-tracking algorithm. The algorithm must check each differential value whether it (potentially) violates DMDP or not. Checking can be

performed "independently" with regard to all the possible differences that have already appeared in MDT. Even if the constraints are not independent, we can over-estimate the situation and we may assume they are so. Specifically, we are trying to estimate the number of $m \times t$ modular sonar sequences that can be constructed from a given $m \times (t-1)$ modular sonar sequence by adjoining one last symbol. It would be equal to the size of the set $S(m,t)$ of GOOD symbols for $f(t)$. Thus, the goal is to estimate or to give some bound on $|S(m,t)|$ when we are given an $m \times (t-1)$ modular sonar sequence.

We will make some reasonable assumptions under which we could recursively estimate the number $N(m,t)$ of $m \times t$ modular sonar sequences as the number $N(m,t-1)$ of $m \times (t-1)$ sequences times $|S(m,t)|$.

Obviously, there are $N(m,1) = m$ sequences of size $m \times 1$. For $2 \le t \le m+1$, we need to estimate the fraction $p(m,t) = |S(m,t)|/m$. We claim that

$$p(m,t) \approx \prod_{h=1}^{\lfloor \frac{t}{2} \rfloor} q(m,t,h), \quad \text{for } 2 \le t \le m+1, \tag{4}$$

where

$$q(m,t,h) \approx 1 - \frac{m(t-2h)}{(m-h+1)^2}, \quad \text{for } 1 \le h \le \lfloor \tfrac{t}{2} \rfloor. \tag{5}$$

From the value given in (4), we may obtain the following:

$$\begin{aligned} N(m,1) &= m, \\ N(m,n) &\approx N(m,n-1)mp(m,n) \\ &= m^n \prod_{t=1}^{n} p(m,t), \quad 2 \le n \le m+1. \end{aligned} \tag{6}$$

The key to the derivation of (6) is to identify the quantities $p(m,t)$ and $q(m,t,h)$ as certain probabilities of related models which simulate the back-track algorithm of the search. The probability model in reality must consist of a set of events, each of whose probabilities are heavily inter-dependent with one another. To make things simple, we use three assumptions discussed below so that the dependence disappears, and the result becomes a simple multiplication of individual probabilities. In doing so, we will adjust a bit further so that the approximation becomes reasonably meaningful. The value $p(m,t)$ will be identified with the probability that an arbitrarily selected symbol at position t satisfies DMDP with respect to all the previous entries of the MDT constructed so far. The first assumption is given as follows:

Assumption 1. For any m and t, the value $p(m,t)$ remains the same no matter which $m \times (t-1)$ modular sonar sequence might be given.

Taking Assumption 1 into account, we will have a sequence of length t with probability $p(m,t)$ given ANY sequence of length $t-1$. The second assumption enables us to factor $p(m,t)$ as a product of some individual probabilities:

Assumption 2. In figuring out the size $|S(m,t)|$, the number of constraints (DMDP) is independent with the depth parameter h (in the definition of DMDP) when a suitable range for h is taken into consideration.

Following Assumption 2, $p(m,t)$ is a product of probabilities of individual events related to the depth parameter h of a given MDT. Note that the RHS of (4) is the product of $q(m,t,h)$'s in the range of h from 1 to $\lfloor \frac{t}{2} \rfloor$. The value $q(m,t,h)$ will be identified with the probability that an arbitrarily selected symbol at position t satisfies DMDP of level h deep of the entries of the MDT constructed so far. This quantity is still too complicated to calculate exactly, and we need the following third assumption:

Assumption 3. The probability $q(m,t,h)$ can be approximated as the conditional probability that an arbitrarily selected symbol at position t satisfies DMDP of level h deep with regard to the entries of the MDT constructed so far, given the condition that it satisfies all the DMDP of level $k < h$ deep.

Under Assumption 3, the value $q(m,t,h)$ can be approximated as the conditional probability that there is no j such that $h < j < t$ and $f(t) - f(t-h) = f(j) - f(j-h) \pmod m$, given the condition that, for each and every k with $1 \le k < h$, there is no j such that $k < j < t$ and $f(t) - f(t-k) = f(j) - f(j-k) \pmod m$. To find this conditional probability and show that it is given as in RHS of (5), we claim that the complementary event has an approximated probability

$$1 - q(m,t,h) \approx \frac{(t-2h)}{(m-h+1)} \cdot \frac{m}{(m-h+1)}. \tag{7}$$

We may easily determine an upper and lower bound on $1 - q(m,t,h)$ which is the fraction of symbols that violates DMDP. Recall that the current position is t, and we are given a sequence of length $t-1$ and the corresponding MDT of depth $t-2$. There are $t-(h+1)$ symbols which violate DMDP for fixed h. It is just the number of entries of MDT at level h deep. Thus, at most $\frac{t-(h+1)}{m}$ of A_m will be BAD for $f(t)$ from row h of MDT, and hence, $\frac{t-(h+1)}{m}$ is an upper bound on $1 - q(m,t,h)$. When we use Fact 1 and Assumption 3, we see that there are at least $\frac{t-(h+1)-(h-1)}{m-(h-1)} = \frac{t-2h}{m-h+1}$ of A_m, which will be BAD for $f(t)$, since $h-1$ symbols (shaded area of the row h in Fig. 1, for example, and using Fact 1) have already been taken care of with regard to the DMDP of level $k < h$ deep. Thus, $\frac{t-2h}{m-h+1}$ is a lower bound on $1 - q(m,t,h)$. Therefore, we have

$$\frac{t-2h}{m-h+1} \le 1 - q(m,t,h) \le \frac{t-(h+1)}{m}.$$

By carefully examining the situation further, we have chosen a factor as shown in (7) and obtained the result given in (5).

In order to check the validity of the estimated number $N(m,n)$ of $m \times n$ modular sonar sequences, we have done a second round search for the values $N(m,n)$ for m up to 14 and n up to $m+1$. These are shown in Table 2 with the calculated number from (5). Further, this relation is plotted in Fig. 2.

Table 2. Comparison of the true and estimated values of $N(m, n)$

m	n	Search	Estimate	m	n	Search	Estimate	m	n	Search	Estimate
7	4	5	5.00	10	9	707	895	13	5	100	100
	5	16	16.1		10	63	103		6	729	738
	6	27	29.5		11	2	0.857		7	3712	3842
	7	16	17.7	11	4	9	9.00		8	12433	13492
	8	2	1.73		5	64	64.1		9	22983	26184
8	4	10	10.5		6	343	350		10	20198	25321
	5	43	43.9		7	1152	1215		11	5922	8835
	6	108	118		8	2209	2479		12	481	852
	7	128	140		9	1857	2190		13	22	11.5
	8	50	54.3		10	533	670		14	1	0.0073
	9	2	2.26		11	35	37.1	14	4	26	26.0
9	4	9	9.33		12	1	0.142		5	262	262
	5	48	48.1	12	4	27	27.5		6	2160	2188
	6	167	173		5	222	223		7	12896	13362
	7	292	326		6	1399	1430		8	53373	57579
	8	249	271		7	5848	6187		9	130547	147892
	9	37	54.2		8	15324	17022		10	168576	209626
	10	3	1.12		9	20155	23392		11	87718	127056
10	4	18	18.0		10	10199	13865		12	14775	27624
	5	110	110		11	1351	2297		13	615	1362
	6	480	499		12	25	67.7		14	2	9.14
	7	1216	1325		13	4	0.0996		15	0	0.0022
	8	1619	1845	13	4	11	11.0				

Fig. 2 shows that the estimated value of $N(m, n)$ fits well with the true value. So, we can say estimation function shows the value's tendency similarly. That is, the tendency of Fig. 3 can be a partial explanation of the decaying tendency of modular sonar sequences.

Remark 1. Estimated values in Table 2 and two figures represent fractions of the value given in (6) divided by $m^2\phi(m)$, since there are at most $m^2\phi(m)$ equivalent but possibly distinct modular sonar sequences on A_m with very high probability. Exception occurs when $g = f$ in Eq.(3) although $u \neq 1$. This event occurs rarely even at small m. We ignored the exceptions.

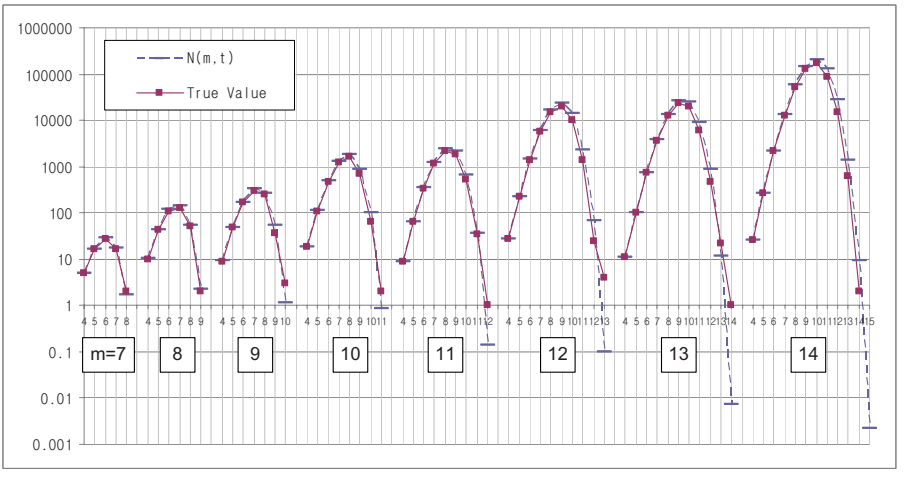

Fig. 2. Comparison of the true and estimated values of $N(m,t)$. From left to right, m runs from 7 to 14, and t runs from 4 to $m+1$ in each group.

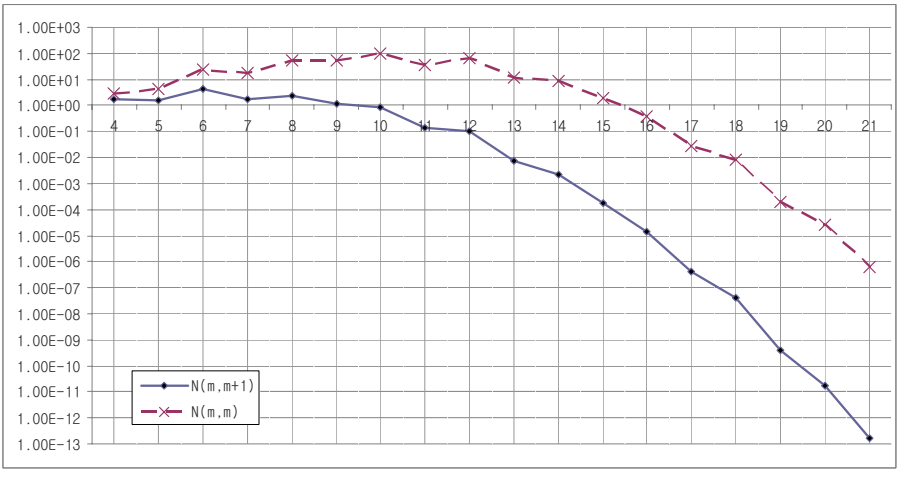

Fig. 3. Estimated values of $N(m,m+1)$ and $N(m,m)$ for some small m

Note that $\frac{N(25,22)}{25^2\phi(25)} \fallingdotseq 20.1$ and $\frac{N(25,23)}{25^2\phi(25)} \fallingdotseq 0.003$. We already know that there is no 25×23 modular sonar sequence from Table 1.

4 Conclusion

We have checked the existence of $m \times n$ modular sonar sequences by computer search for some small values of m, and estimated the number of inequivalent

examples for various values of m by carefully examining the back-track algorithm for the search.

From this estimate, we could have concluded that no full-size modular sonar sequence exists for m beyond a certain value. This is, however, not true, since there are some algebraic constructions which give full size examples (of length $m+1$ on m symbols) for infinite values of m. We could safely guess that any full-size example for large values of m must be either from an algebraic construction, or else the probability that it exists is extremely small.

We still leave the following problems unsolved:

Unsolved Problem 1. Find an example of 35×35 modular sonar sequences (mod 35) or prove that none exists.

Unsolved Problem 2. Generalize the above to the case of $m = p(p+2)$ being a product of twin primes.

Unsolved Problem 3. Find infinitely many values of m for which an $m \times (m+1)$ modular sonar sequences do not exist.

Unsolved Problem 4. Except for m being a prime or one less than a prime power, would the fact that the value in (6) is close to zero imply non-existence?

Unsolved Problem 5. How accurate is the estimate in (6)?

Unsolved Problem 6. Could a similar approach be used to estimate the number of Costas arrays, see [7] ?

References

1. Costas, J.P.: Medium constraints on sonar design and performance. FASCON CONV. Rec. 68A–68L (1975)
2. Erdös, P., Graham, R.L., Ruzsa, I.Z., Taylor, H.: Bounds for arrays of dots with distinct slopes or lengths. Combinatorica 12, 1–6 (1992)
3. Moreno, O., Games, R.A., Taylor, H.: Sonar sequences from costas arrays and the best known sonar sequences with up to 100 symbols. IEEE Trans. Inform. Theory 39(6) (November 1987)
4. Gagliardi, R., Robbins, J., Taylor, H.: Acquisition sequences in PPM communications. IEEE Trans. Inform. Theory IT-33, 738–744 (1987)
5. Games, R.A.: An algebraic construction of sonar sequences using M-sequences. SIAM J. Algebraic Discrete Methods 8, 753–761 (1987)
6. Golomb, S.W., Gong, G.: Signal Design for Good Correlation. Cambridge University Press, Cambridge (2005)
7. Golomb, S.W., Taylor, H.: Two-dimensional synchronization patterns for minimum ambiguity. IEEE Trans. Inform. Theory IT-28, 263–272 (1982)
8. Yoon., S.-J., Song, H.-Y.: Existence of Modular Sonar Sequences of Twin-Prime Product Length. In: Golomb, S.W., Gong, G., Helleseth, T., Song, H.-Y. (eds.) SSC 2007. LNCS, vol. 4893, pp. 184–191. Springer, Heidelberg (2007)
9. Silverman, J., Vickers, V.E., Mooney, J.M.: On the number of Costas arrays as a function of array size. Proceedings of the IEEE 76(7) (July 1988)