Turk J Elec Engin, VOL.15, NO.1 2007, © TÜBİTAK

Data Mining and Costas Arrays

Konstantinos DRAKAKIS*
Claude Shanon Institute and Electronic and Electrical Engineering University College Dublin-IRELAND
e-mail: Konstantinos.Drakakis@ucd.ie

Abstract

Costas arrays are used in RADAR (and SONAR) engineering to represent frequency-hopping patterns that optimize the RADAR's performance. In this work, using all available Costas arrays up to and including order 26, as well as data mining techniques, we investigate how the Costas property and the balance of signs in the difference triangle of a permutation are related. Our conclusion is that there is sufficient evidence to believe that the mechanism responsible for the formation of Costas arrays changes from low to high orders.

1. Introduction

In this work we will see a novel approach to study Costas arrays based on data mining techniques. We will begin by offering the definition and fundamental properties of Costas arrays, as it is then easier to see what they can be useful for; subsequently, we will offer a summary of what we know today about them. Please note that, although the term "matrix" is much more common today in all branches of Mathematics, it is still customary to talk about "Costas arrays" instead of "Costas matrices", perhaps because a Costas array is used in RADAR engineering to represent an arrangement of frequencies in time; we will talk more about applications of Costas arrays below.

1.1. Definition

A Costas array is a special case of a permutation matrix, i.e., a square matrix whose elements are equal either to 0 or 1, and which contains exactly one element equal to 1 per row and column. The additional property needed, which defines Costas arrays, is that all vectors between 1s are different.

In more precise language:

Definition 1 *Let* $n \in \mathbb{N}^*$*, and let* A *be a permutation matrix; if* $a_{ij} = 1$*, set* $f(j) = i$*, with* $i, j \in \{1, 2, \ldots, n\}$*. We call* A *a* Costas array *(of* dimension *or* order n*) if and only if the following condition is satisfied:* $\forall i_1, i_2, i_3, i_4 \in \{1, 2, \ldots, n\}, i_1 \leq i_2, i_3 \leq i_4 : (i_1 - i_2, f(i_1) - f(i_2)) = (i_3 - i_4, f(i_3) - f(i_4)) \implies i_1 = i_3, i_2 = i_4$*. In other words,* A *is a Costas array if and only if all vectors of the form* $(i_1 - i_2, f(i_1) - f(i_2)),\ i_1, i_2 \in \{1, 2, \ldots, n\}, i_1 < i_2$*, i.e., vectors between 1s, are distinct. For shorthand, we will occasionally write* $f = (f(i))$*, when we want to show the elements of* f*.*

*The author holds a Diploma in Electrical and Computer Engineering from NTUA, Athens, Greece, and a Ph.D. in Applied and Computational Mathematics from Princeton University, NJ, USA.

Here, $f(j) = i$ expresses the fact that the element of column j that is equal to 1 lies at the ith position of the column. Observe that, since each column and row has a unique element equal to 1, the others being 0, f is a bijection, hence f^{-1} is a function, too: $f^{-1}(i) = j$. Note that f characterizes A unambiguously.

A direct consequence of this definition is that Costas arrays come in sets of 4 or 8:

Definition 2 *Let the Costas array $A = [a_{ij}]$ be of order $n \in \mathbb{N}^*$ and correspond to the permutation $f = (f(i))$. The* vertical flip *of A is the matrix $A_v = [a_{n+1-i,j}]$ corresponding to $f_v = (n + 1 - f(i))$, while the* horizontal flip *of A is the matrix $A_h = [a_{i,n+1-j}]$ corresponding to $f_h = (f(n + 1 - i))$. We take everywhere $i \in \{1, 2, \dots, n\}$.*

If we start with a Costas array and flip it either horizontally, or vertically, or along the main diagonal, the resulting matrix is still Costas. These 3 transformations produce 8 matrices in total, or 4 if the initial Costas array happens to be symmetric, so that flipping it around the main diagonal leaves it unchanged.

1.2. Difference Triangle

The definition of a Costas array can be rephrased to make it easier to visualize and grasp, by collecting vectors together according to their first coordinate. The vectors at hand are as many as the possible choices of $i_1, i_2 \in \{1, 2, \dots, n\}$ with $i_1 < i_2$, i.e., $\frac{n(n-1)}{2}$ in total, and, out of these vectors, exactly $n - k$ have their first coordinate equal to k, for $k = 1, 2, \dots, n$ (to be specific, it is those vectors that correspond to $i_1 = i,\ i_2 = i + k$ for $i = 1, 2, \dots, n - k$); these vectors can be collected in a set, say S_k. Within each S_k a vector can be represented only by its second coordinate, as the first is the same for all members of this set.

Consider two vectors $v_1 \in S_{k_1}$ and $v_2 \in S_{k_2}$. In the context of the definition of a Costas array, we need not worry whether the two vectors are equal if $k_1 \neq k_2$, and if $k_1 = k_2$ we only need to check the second coordinate of the vectors to make sure. So, we can list the second coordinates of the vectors of S_k in a row, and make sure that within this row no number appears twice. If we order the rows one on top of the other, left adjusted or centered, the row corresponding to S_1 being the topmost, and the row corresponding to S_{n-1} being the bottommost, we will obtain a triangular structure: we call this the *difference triangle* of a permutation matrix.

Definition 3 *Let A be a permutation matrix. The difference triangle of A, T or $T(A)$, when we want A to appear explicitly, is a triangular structure of $n - 1$ rows that has the entries $t_{ij} = f(j) - f(j + i)$, $i = 1, 2, \dots, n - 1$, $j = 1, 2, \dots, n - i$.*

The previous analysis proves:

Theorem 1 *Let A be a permutation matrix. It is a Costas array if and only if no number appears twice in a row of $T(A)$.*

1.3. Applications of Costas Arrays

Costas arrays have many applications in SONAR and RADAR Engineering [4, 6, 8, 9], which is where they originated from. Consider a RADAR that can transmit pulses at n different frequencies; how should we then

Table 1. Number of Costas arrays per order found by exhaustive search

Order	Number	Order	Number	Order	Number
1	1	10	2160	19	10240
2	2	11	4368	20	6464
3	4	12	7852	21	3536
4	12	13	12828	22	2052
5	40	14	12752	23	872
6	116	15	19612	24	200
7	200	16	21104	25	88
8	444	17	18276	26	56
9	760	18	15096	27	?

arrange the frequencies into a time-frequency transmission pattern so as to optimize RADAR performance? An energy argument [8, 9] shows that for optimal detection the energy at a given time should not be split in two or more frequencies, nor should the energy at a given frequency be split in two or more pulses disjoint in time. Therefore, each frequency should be used for exactly one pulse, and this turns the time-frequency pattern into a (square) permutation matrix of order n. Among those matrices, are some better than others?

Consider operation in a noiseless environment, where the object of the RADAR is to determine both the position and the velocity of a target. Then, the pattern reflected by the target and received by the RADAR is identical to the transmitted one, except for being translated in time and frequency: the time delay shows the distance of the target, and the frequency shift its velocity through the Doppler effect. To determine the two shifts, the cross-correlation between the transmitted and the received pattern is computed, and the correct shifts clearly correspond to the position of the maximum, which should for this reason be as large as possible relatively to the side-lobes of the cross-correlation, especially when noise is present, as it always is in actual applications.

This is precisely where the Costas property is useful: if the permutation matrix satisfies the Costas property, no two (time-frequency) vectors between two pulses are the same, and hence the cross-correlation side-lobes can never be larger than 1!

1.4. Summary of Current Results

Algorithms have been proposed for the construction of Costas arrays (the Welch, Golomb, and Taylor constructions, along with their variants) [5, 6], which work when the order of the matrix is a bit lower than a prime or a power of a prime, or sometimes equal to that. Moreover, recent advances in computers have made possible the determination of all Costas arrays of orders up to and including 26 [3, 7] through exhaustive search, but the exponential increase of complexity (there are $n!$ permutation matrices of order n) does not allow us to be very optimistic that higher orders will be tackled soon. The number of Costas arrays per order found by extensive search is given in Table 1.

With the exception of the advances in exhaustive search, and of a method proposed recently [4], no other widely applicable method for the discovery of new Costas arrays has been found in the last 20 years. This recent method builds larger Costas arrays out of Golomb and Welch constructed ones, and it yielded 4 (32 with the symmetries) previously unknown Costas arrays of orders less than 100. Despite its limited results, it is important because it signified a new approach, that of a “clever brute force”: in other words, the computationally expensive exhaustive search of exponential complexity and the theoretically proved properties of the mathematical construction algorithms were combined into a hybrid algorithm.

Before we can prove properties of Costas arrays, though, we need to positively determine some plausible ones, and the most efficient way to do this is through the study of the Costas arrays we know so far, that is through data mining.

In particular, we do *not* know the following:

- Are there Costas arrays for any order?
- Are there Costas arrays for orders 32 and 33, which are the lowest orders for which Costas arrays have not yet been found [6, 4]?

2. Data Mining

It is the direction of "clever brute force" that we wish to pursue in this work. What we propose to do is to use the results of the exhaustive search to investigate for patterns in Costas arrays, and thus make the search of new arrays easier. We will focus on two features of Costas arrays:

1. The value of $f(1)$, i.e., the starting number of the permutation corresponding to the Costas array.
2. The signs of the elements of the difference triangle.

The basis of our reasoning in both cases is a property of the difference triangle:

Theorem 2 *Let A be a permutation matrix of order n. Then, $T(A)$ contains exactly $n-i$ elements equal to i in absolute value, $i = 1, 2, \ldots, n-1$.*

Proof It is the case that $\{f(1), f(2), \ldots, f(n)\} = \{1, 2, \ldots, n\} = S_n$, as f is a bijection from S_n to itself. This implies that $[|f(i) - f(j)| : i, j = 1, 2, \ldots, n,\ i < j] = [j - i : i, j = 1, 2, \ldots, n,\ i < j]$, where we use square brackets to denote *multisets*, i.e., sets that can contain the same value more than once. We can see that $j - i = k \Leftrightarrow j = k + i$, and since we want $1 \leq i < k + i \leq n$ the value k appears for $i = 1, 2, \ldots, n-k$, i.e., $n-k$ times. This completes the proof. $\square$

We will proceed now to take as a working assumption that the entries of the difference triangle are reasonably uncorrelated (a more precise description of the correlation of the entries of the difference triangle is given in [1]).

Definition 4 *Let A be a Costas array, let $T(A)$ be its DT, and suppose that $T(A)$ contains p positive entries and q negative. Then, the* sign balance *of A is $q - \dfrac{p+q}{2} = \dfrac{q-p}{2}$, and the closer this number is to 0, the better balanced A is.*

Conjecture 1 *The better balanced the numbers of positive and negative entries are in the difference triangle, the more probable it is that the permutation at hand corresponds to a Costas array.*

This seems reasonable: start building the difference triangle of a permutation by randomly assigning positive values only to its entries, so that the number of entries of a particular value satisfies Theorem 2. Once this procedure is complete, there will probably be entries of the same value in a particular row. Choose a pair of equal entries and change the sign of one of them: the entries are no longer equal. The closer the numbers of positive and negative signs we can assign are, the more the pairs of equal values we can turn into opposite values will be, and the more the rows of the difference triangle will be free of repetitions. Notice here that we deliberately ignore the issue of realizability of the difference triangle we built: it may correspond to no permutation at all. Our assumption that the entries of the difference triangle are "reasonably uncorrelated" implies that this won't be a serious problem, and that we intend to ignore this issue in our approach.

Before we test this conjecture, we will test a simple derivative of it.

2.1. Starting Numbers

Given that A is a Costas array of order n, and that it corresponds to the permutation $f(1)f(2)\dots f(n)$, what is the probability $\mathbb{P}[f(1)=i]$, $i=1,2,\dots,n$? If we knew that Costas arrays "prefer" a particular $f(1)$ consistently in all orders, we could make it easier to search for Costas arrays of an order n where none has been found yet, like 32 and 33 for example, as we would reduce the complexity of the search from $n!$ to $(n-1)!$, which is a considerable reduction. Further analysis, such as determining $\mathbb{P}[f(2)=j|f(1)=i]$ would allow us to reduce the complexity of the search even further, but here we will focus on the more basic statistics of the starting number only:

Conjecture 2 *Consider the probability distribution $\mathbb{P}(f(1))=i$, $i=1,2,\dots,n$, where f is a Costas permutation of order n. Then its maximum, or at least a large value, occurs for $\left\lceil \frac{n}{2} \right\rceil$, i.e., the mid-number.*

The first entry of row i of the difference triangle is $f(1)-f(i+1)$, $i=1,2,\dots,n-1$. If we want the signs of these entries to be as balanced as possible, we need to choose $f(1)$ as Conjecture 2 dictates. Therefore, Conjecture 2 is consistent with Conjecture 1.

Figure 1 shows the histograms of $f(1)$ for the Costas arrays of some orders up to and including 25: all histograms up to order about 20 look very nicely centered, just like the first 3 graphs in the figure, and validate Conjecture 2. For larger orders, though, a rather erratic behavior settles in, as the last 3 graphs, show: even those histograms compatible with Conjecture 2 (as in orders 23 and 24) show many strong local maxima, while the last graph, for order 25, is definitely incompatible with it.

The deviation in large orders from the rule onjecture 2 sets may be due to the small number of Costas arrays in these orders (see Table 1), although for small orders the number of Costas arrays is also small and we do not observe a similar deviation. It might also be due to a radical change in the mechanism that produces Costas arrays of larger orders. In any case, based on this evidence we cannot conclude that the mid-number is the best guess for $f(1)$ if we wish to find a new Costas array of order 27, 28 etc.

2.2. Balance of Signs in the Difference Triangle

Since the test for the value of $f(1)$ proved rather inconclusive, we will proceed to check here the full Conjecture 1: if it turns out to be true, it will be efficient to look for Costas arrays primarily among the permutations with a difference triangle with balanced positive and negative signs, if we assume that we have an efficient algorithm to generate such permutations easily in the first place.

The difference triangle carries much more information than it is necessary in order to specify a permutation; to see this, observe that even the first row alone would suffice for this purpose. We will

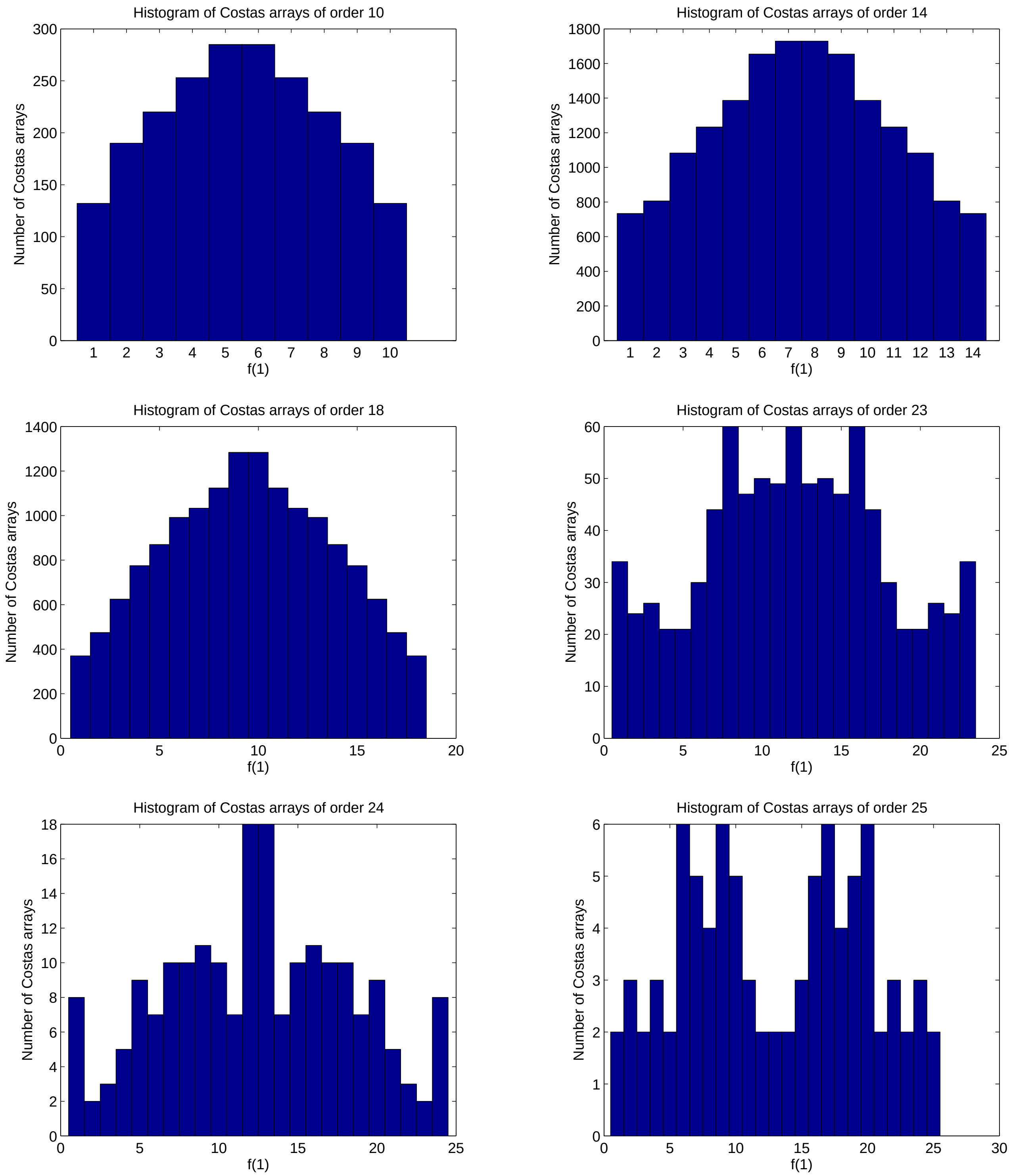

Figure 1. Histograms of $f(1)$ for the Costas arrays of various orders

demonstrate now that even if all entries are wiped out entirely, and only their signs are kept, the permutation is still completely specified (see [10]). Consider, for example, the difference triangle:

$$\begin{array}{l} +--+- \\ --++ \\ --+ \\ +- \\ - \end{array}$$

It actually corresponds to a Costas array, but this is not important for the purposes of this example. How can we recover the permutation it expresses?

- First of all, the order is 6, as the first row has 5 elements: the possible entries are $1, 2, 3, 4, 5, 6$.
- By looking at the leftmost column, we see that $f(1)$ is larger than 2 elements and smaller than 3; the only possibility is $f(1) = 3$.
- $f(2)$ is less than all remaining entries (the second column is all negative), so $f(2) = 1$.
- The third column indicates that the third element is smaller than only one of the remaining entries, so $f(3) = 5$.
- $f(4)$ is greater than the remaining entries, so $f(4) = 6$, as the fourth column is all positive.
- $f(5)$ is less than $f(6)$, as the fifth column is negative, so $f(5) = 2$ and $f(6) = 4$.

This leads to the following simple algorithm for the determination of a permutation from the signs of its difference triangle:

Algorithm

1. The order of the permutation is one more than the length of the first row; denote it by n.
2. Set $A = \{1, 2, \ldots, n\}$ to be the set of available numbers; set $i = 0$;
3. Increase i by one; count the number j of positive signs in column i of the triangle, and choose the $j+1$ largest number (the number larger than j others), say k, among those in S; set $f(i) = k$ and remove k from S.
4. If $i < n$ repeat the previous step, otherwise stop.

This is a relatively simple inverse problem in permutations, and many others can be formulated: we could, for example, discard the order of the elements within the rows, or discard the signs and keep the absolute values only, or even do both, and investigate whether reconstruction is still possible. We will not pursue such questions further in this work, however.

The difference triangle of a permutation of order n has $\dfrac{n(n-1)}{2}$ entries, so it can have $0 \leq k \leq \dfrac{n(n-1)}{2}$ negative entries; the best balanced triangles will then have $\left\lceil \dfrac{n(n-1)}{4} \right\rceil$ entries. We could follow the same steps we followed in the previous experiment, but there is an additional complication that must be tackled here: whereas there are exactly $(n-1)!$ permutations with a given $f(1)$ no matter what this

$f(1)$ is, this is no longer true when we categorize permutations according to k, because some categories contain many more permutations than others. As an example, only one permutation corresponds to $k=0$, namely $n, n-1, \ldots, 1$. If then we use the same type of histogram as before, measuring the number of Costas arrays corresponding to a given k, which henceforth we will denote by $D(n,k)$, we may obtain spurious results: a histogram may be showing a large $D(n,k)$ just because there may be many permutations with this k to begin with, and not because this value of k somehow improves the chances for a Costas array. Therefore, we should present a chart of $D(n,k)/C(n,k)$, where $C(n,k)$ denotes the total number of permutations corresponding at k; we will call this fraction the *concentration* of Costas arrays. A value of the concentration well above 1 indicates that the corresponding k is a good one to look for Costas arrays.

This leads naturally to the need to count $C(n,k)$; this result is standard in Combinatorics and it can be found in [10], for example, but we will derive it here, too: if we take a permutation contributing to $C(n,k)$ and we place $n+1$ in it at position i from the left, we will create a new permutation of order $n+1$ contributing to $C(n+1, k+i-1)$. We establish thus a recursion: if we consider a permutation contributing to $C(n,k)$ and we remove n completely from it, we end up with a permutation contributing to $C(n-1, k-(i-1))$ assuming n was at position i from the left. Hence:

$$C(n,k) = \sum_{i=1}^{n} C(n-1, k-(i-1))$$

with the boundary conditions $C(n,k)=0,\ k<0$ or $k > \dfrac{n(n-1)}{2}$, and $C(2,0)=C(2,1)=1$. If we further define the generating function $F_n(z) = \sum_{k=0}^{\frac{n(n-1)}{2}} C(n,k) z^k$, we can turn the relation above into:

$$F_n(z) = F_{n-1}(z)(1+z+\ldots+z^{n-1})$$

with $F_2(z) = 1+z$. This recursion allows us to compute $C(n,k)$ as the coefficient of z^k in $F_n(z)$.

Figure 2 shows the concentration of Costas arrays at the same orders as Figure 1. The concentration is plotted not against k, but rather against the excess of negative signs $k - \left\lceil \dfrac{n(n-1)}{4} \right\rceil$, that we defined as the measure of balance in Definition 4, so that 0 corresponds to perfect balance. The conclusions are similar to the previous experiment: up to order about 20 the concentration seems indeed quite high around 0, confirming Conjecture 1; but as we get close to 25 pretty unbalanced Costas arrays start appearing (at 23), until they completely dominate the picture at 25. Following this evidence, we cannot conclude that the Conjecture 1 holds for all orders, and once more it appears that a completely different mechanism is responsible for the formation of Costas arrays of higher orders.

3. Conclusion

Costas arrays of order up to about 20 show an extraordinary uniformity, as most of them appear to be formed in accordance with the principle that positive and negative entries in the difference triangle are balanced, i.e., approximately equal in number. This uniformity breaks down, though, as we start getting closer to order 25, currently the highest order for which we know all Costas arrays. It is then possible that for higher orders

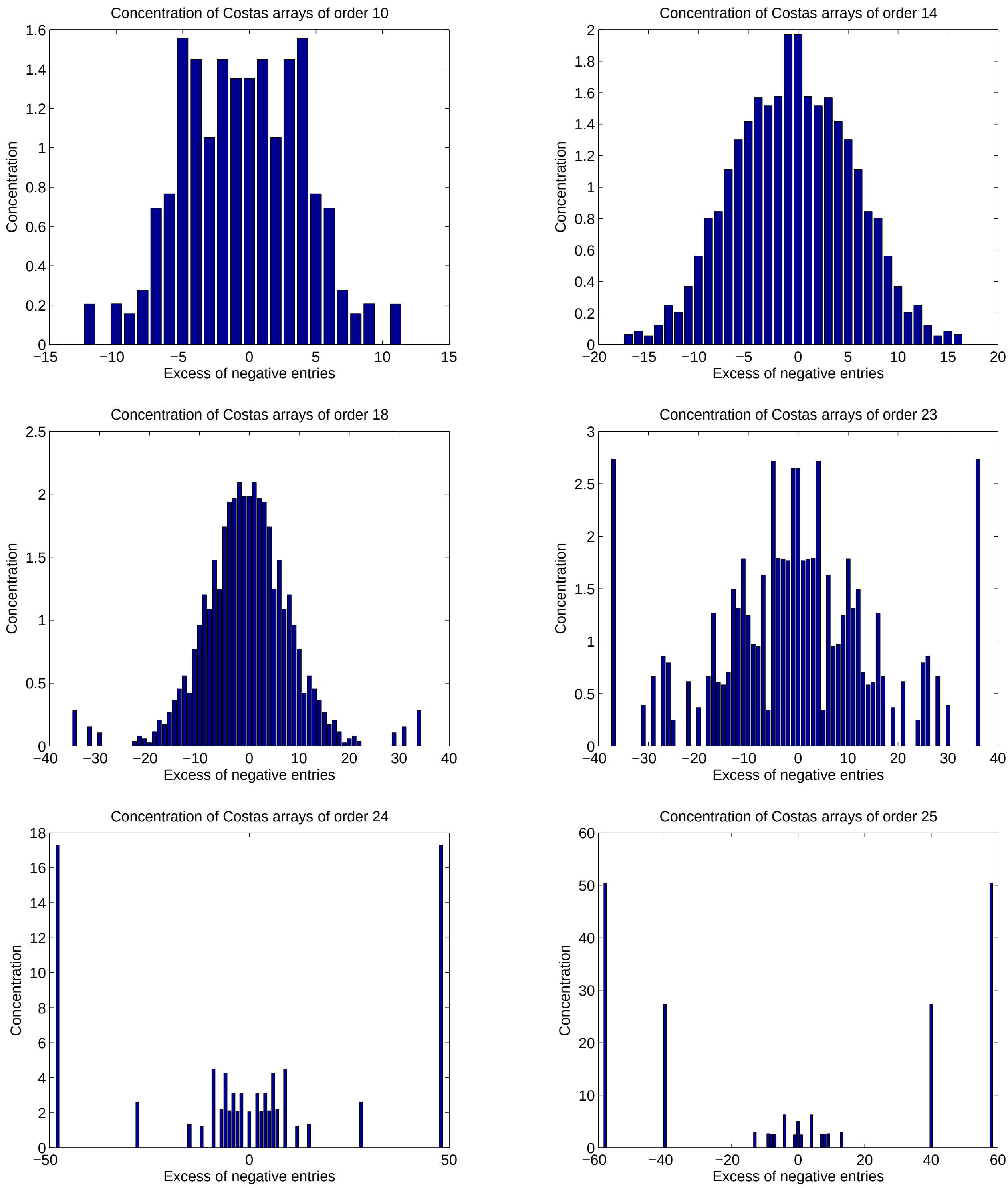

Figure 2. Concentration of Costas arrays of various orders

completely different mechanisms take over. This is a reasonable assumption: Conjecture 1 is oversimplifying as it is not taking into account the values of the entries of the difference triangle, but only their sign instead. As the order grows, the defining constraints of the Costas array become tighter and tighter (there are $O(n^3)$ constraints for n integers, if the order is n [1]), and it seems possible that from a point onwards the analysis carried out with signs alone is not enough.

Acknowledgements

The author would like to thank Dr. Scott Rickard of the Department of Electrical Engineering of University College of Dublin for introducing him to the problem, and providing him with the data used in this work, as well as some papers [7, 4]. He would also like to thank the anonymous referees of this paper who contributed to its improvement it through their comments and constructive criticism.

References

[1] J. Silverman, V. E. Vickers, J. M. Mooney, *On the number of Costas arrays as a number of array size*, Proceedings of the IEEE, Vol. 76, No. 7, July 1988.

[2] D. Huw Davies, *On the Density of Costas Arrays*, IEEE Trans. Information Theory, October 1989.

[3] J. K. Beard, J. C. Russo, K. Erickson, M. Monteleone, M. Wright, *Combinatoric Collaboration on Costas Arrays and Radar Applications*, IEEE Radar Conference, Philadelphia, PA, USA (2004).

[4] S. Rickard, *Searching for Costas arrays using periodicity properties*, IMA International Conference on Mathematics in Signal Processing, The Royal Agricultural College, Cirencester, December 2004.

[5] S. W. Golomb, *Algebraic Constructions for Costas Arrays*, Journal of Combinatorial Theory, Series A 37, 13–21 (1984).

[6] S. W. Golomb, H. Taylor, *Constructions and Properties of Costas Arrays*, Proceedings of the IEEE, Vol. 72, No.9, September 1984.

[7] J. K. Beard, *Combinatoric Collaboration on Costas Arrays and Radar Applications*, Slide presentation in RADARCON (2004).

[8] J. P. Costas, *Medium constrains on sonar design and performance*, Technical Reposrt Class 1 Rep. R65EMH33, GE Co.

[9] J. P. Costas, *A study of detection waveforms having nearly ideal range-doppler ambiguity properties*, Slide presentation in IEEE Radar Conference, Philadelphia, PA, USA (2004).

[10] R. P. Stanley, *Enumerative Combinatorics, Vol. 1*, Cambridge 1997.