

A Computing Research Example: Following a Question

Ronald I. Frank
rfrank@Pace.edu

Abstract

We give an example of research: how an observation and a question about what it means leads to new insights - if diligently followed. This is an ongoing example, still in progress and requires only basic college mathematics to follow. We start from an observation about cubic arrays and ask what the formula implies. Following the trail, we discover the possibility of defining "arrays" that have strange properties such as negative length and fractional dimension. Along the way, we discover a polynomial that generates all of the sub arrays of a given array. This is a problem of interest in the design of parallel computers such as hypercube machines. It also pops up in data mart data cubes used in OLAP. We talk in terms of data arrays, but the reader can also think of arrays of processors or of array structured clusters.

Keywords: Research example, Array structure, Characteristic polynomial, Fractional dimension, Complex Length, Negative Length, Array function.

1. INTRODUCTION

When teaching research, we often tell students to be curious (in the active sense) and find interesting questions, then follow one to its conclusion. I had two observations and asked myself what do they mean? The following are a few of the results that followed and some left over questions and new conjectures. Much of the basic results discussed here come from (Frank 2002) and the three references (Frank 2003). The central result is the existence of a characteristic polynomial for the structure of a regular array. See section 4 below for a worked example.

Although the focus here is as on an example of following a question in research, this turns out to be also an update report of the four basic references. The report here has been much streamlined, the dual view notation is new and improved, the fractional dimension result is new, the presentations of negative length, complex length, the conjectures, and the note on topology are also new. However, none of the new results are rigorously derived here. They are only reported as part of the research example. The full developments are being submitted elsewhere for publication.

2. ORIGINAL QUESTION(S)

First, given a data cube of N dimensions with shape vector $\{n, n, \dots, n\}$, it contains n^N elements. This formula "represents" the data array. What comes of this?

Second, any power form immediately suggests the Binomial Theorem or Series, so what comes of that?

3. BINOMIAL THEOREM

[Knuth 1997 pp. 57 and following]. See Appendix B on the Binomial Coefficients. We will use the symbol T_n^N to mean a representation of an array of length n and dimension N . Applying the binomial expansion, we get.

$$T_n^N \approx (n)^N = (n-1+1)^N = ((n-1)+1)^N \\ = \sum_{k=0}^N \binom{N}{k} (n-1)^k (1)^{N-k}$$

$$T_n^N \approx \sum_{k=0}^N \binom{N}{k} (n-1)^k \approx \sum_{k=0}^N \binom{N}{k} T_{n-1}^k \\ \textit{all smaller arrays.}$$

This means that ANY cube can be decomposed into arrays of smaller length and all dimensions less equal the original dimension. This immediately suggested recursion, which is shown in (Frank 2002). The result is that any regular array (cubes or blocks) can be expanded in a polynomial whose powers are null arrays. Later, a simpler derivation was found as shown below

Example [case N = 0].

The shape vector of a scalar (0-D array) is { }, i. e., an empty vector. Its representation is $(0)^0 = T_0^0$

Example [case N = 1].

The shape vector of a 1-D array of length n is {n}. Its representation is $(n)^1 = T_n^1$. In the references it is shown that

$$T_n^1 = (n)T_0^0 + T_0^1$$

where T_0^1 is just {0}, the 1-D null array containing no cells. This is a canonical form for an n-vector array. We interpret this to say that a one-dimensional n-vector array is represented by its n scalar cells and a null vector of dimension 1. This is a new and very different representation of the structure of a one-dimensional finite array.

As a polynomial (monomial) we notice that the length of the array is the negative of the root ($0=n+x$).

$$0 = nT_0^0 + T_0^1$$

4. CREATING A {2, 3} ARRAY

Following along we introduce an idea from linear algebra - the outer product that generates higher dimensional arrays. By taking outer products of 1-D arrays, we can get all of the regular data arrays, cube or block. We show the {2,3} as an example.

$$T_{2,3}^2 = (2T_0^0 + T_0^1) \otimes (3T_0^0 + T_0^1) = 6T_0^0 + 5T_0^1 + T_0^2$$

We have used

$$T_0^i \approx (0)^i \approx \{0, 0, \dots, 0\} \text{ of length } i$$

$$T_0^j \approx (0)^j \approx \{0, 0, \dots, 0\} \text{ of length } j$$

$$T_0^i \otimes T_0^j \approx (0)^i (0)^j = (0)^{i+j} \approx \{0, 0, \dots, 0\} \text{ of length } (i+j) \approx T_0^{i+j}$$

$$[\forall \text{ integer } i, j \geq 0]$$

The meaning of these terms is that they are null arrays, containing NO substructure, but they define the dimensionality of the term they are in.

T_0^0 is a 0-dimensional array containing no scalars, by default.

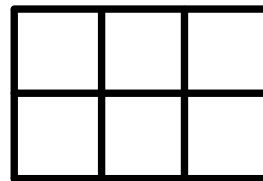
T_0^1 is a 1-dimensional array containing no scalars.

$T_0^2 \equiv T_{0,0}^2$ is a 2-dimensional array containing no scalars, and no vectors.

$T_0^3 \equiv T_{0,0,0}^3$ is a 3-dimensional array containing no scalars, no vectors, and no 2-dimensional arrays. Etc.

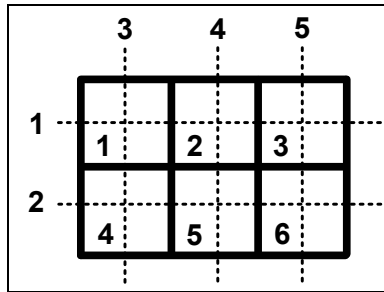
The picture we can use of the array is an "egg crate view"

$$T_{2,3}^2 = (2T_0^0 + T_0^1) \otimes (3T_0^0 + T_0^1) = 6T_0^0 + 5T_0^1 + T_0^2$$



The terms of the array expansion in null arrays can be interpreted as representing the structure of this {2, 3} array.

$T_{2,3}^2 = 6T_0^0 + 5T_0^1 + T_0^2$ means that there are 6 cells (0-dimensional scalars), arranged in 5, one-dimensional vectors, forming one, two-dimensional array. We call this polynomial the array structure characteristic polynomial of the array. This example can be better seen in the next diagram.



By just following an observation, using the binomial expansion, and ordinary outer products we have stumbled across a generating function for any regular array that counts the totality of all sub arrays of any dimension.

Since an array is a geometrical object, its structure is not changed if we interchange the order of its dimensions. This is transposing it in N-dimensions. We note that the natural commutivity of outer product ensures that the array structure characteristic polynomial also does not change. I.e., it is invariant under permutations of the axes.

5. THE GENERAL DISCRETE CASE

(Frank May 2002).

$$T^N_{(n_1, n_2, n_3, \dots, n_{N-2}, n_{N-1}, n_N)} = (n_1 T_0^0 + T_0^1) \otimes (n_2 T_0^0 + T_0^1) \otimes \dots \otimes (n_{N-1} T_0^0 + T_0^1) \otimes (n_N T_0^0 + T_0^1)$$

$$T^N_{(n_1, n_2, n_3, \dots, n_{N-2}, n_{N-1}, n_N)} =$$

$$\sum_{j=0}^N T_0^j \left(\sum_{\binom{N}{N-j}} [(n_{i_1}) \times (n_{i_2}) \times \dots \times (n_{i_{N-j-1}}) \times (n_{i_N})] \right)$$

The coefficients are the count of the number of objects of dimension j making up the array whose shape vector has the coordinates n_k for $k=1, \dots, N$.

The coefficients of the array structure polynomial are symmetric functions (sum over all combinations of shape vector entries taken (N-j) at a time for the jth coefficient). This means that the polynomial is invariant under all permutations of the dimensions. We expect this since a structure representation should not change if we

transpose axes (rotate the array in N dimensions). Again, The lengths of the monomial factor vectors are the negative of the roots of the polynomial. (Frank 2002).

A mystery solved.

The mystery was what is the meaning of the phenomenon that the number of entries in an array is independent of its dimensions? For example we can store only one item in a scalar, a 1-D 1-element vector, a {1,1} matrix, a {1, 1, 1} cube, etc. The explanation is their structure polynomials.

They all start with $(1)T_0^0$ and are of the form $[(1)T_0^0 + T_0^1]^k$, where $k = 0, 1, 2, 3,$ etc.

6. ARRAY CATENATION

"ANTI-GLUE"

Two data arrays can be combined along any axis whose complimentary axes are conformal (Falkoff and Iverson 1970). For example, we can catenate a {2, 1, 4} with a {2, 2, 4} along the second axis generating a {2, 3, 4} array. We use APL "," for catenation.

$$(T_{2,1,4}^3), (T_{2,2,4}^3) = (T_{2,3,4}^3)$$

In terms of their shape vectors we get {2, 1, 4} + {2, 2, 4} = {2, 3, 4}. In terms of their polynomial representations, the factors and the result are:

$$\begin{aligned} (T_{2,1,4}^3) &\approx (2T_0^0 + T_0^1) \otimes (1T_0^0 + T_0^1) \otimes (4T_0^0 + T_0^1) \\ &= 8T_0^0 + 14T_0^1 + 7T_0^2 + T_0^3 \\ (T_{2,2,4}^3) &\approx (2T_0^0 + T_0^1) \otimes (2T_0^0 + T_0^1) \otimes (4T_0^0 + T_0^1) \\ &= 16T_0^0 + 20T_0^1 + 8T_0^2 + T_0^3 \\ (T_{2,3,4}^3) &\approx (2T_0^0 + T_0^1) \otimes (3T_0^0 + T_0^1) \otimes (4T_0^0 + T_0^1) \\ &= 24T_0^0 + 26T_0^1 + 9T_0^2 + T_0^3 \end{aligned}$$

However, the sum of the polynomials of the catenands does not equal the polynomial of the resultant catenation. We cannot simply add catenand polynomials to get the polynomial of the result array.

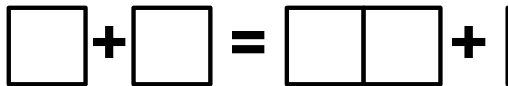
$$\begin{aligned} (T_{2,1,4}^3) &\approx 8T_0^0 + 14T_0^1 + 7T_0^2 + T_0^3 \\ + (T_{2,2,4}^3) &\approx +16T_0^0 + 20T_0^1 + 8T_0^2 + T_0^3 \\ (T_{2,3,4}^3) &\neq 24T_0^0 + 34T_0^1 + 15T_0^2 + 2T_0^3 \\ (T_{2,3,4}^3) &\approx \frac{24T_0^0 + 26T_0^1 + 9T_0^2 + T_0^3}{\text{Difference}} \\ \text{Difference} &= 0T_0^0 + 8T_0^1 + 6T_0^2 + 1T_0^3 \end{aligned}$$

The difference, which must be subtracted from the catenand sum to give the catenation polynomial, is exactly the shape of the catenation interface: This is a general theorem.

$$\begin{aligned} (T_{2,0,4}^3) &\approx (2T_0^0 + T_0^1) \otimes (0T_0^0 + T_0^1) \otimes (4T_0^0 + T_0^1) \\ &= 0T_0^0 + 8T_0^1 + 6T_0^2 + T_0^3 \end{aligned}$$

This is also a property of natural space.

If you take two wooden boxes and put them next to each other and wanted to make a uniform two-compartment box, you would have to take away a side of one of them. Otherwise, the partition would not be like the other walls.



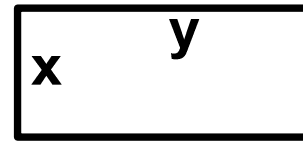
We call the interface subtracted, the "anti-glue" since we subtract it when combining. When breaking up an array by decatenation, the anti-glue must be added back so that the polynomials work out correctly. Similarly, if we wanted to break up the two-compartment box above, we would have to add back the side, which we took away in the catenation.

By applying our array structure characteristic polynomial (ASCP) to a standard array operation, we have discovered a "law" of geometry.

7. CONTINUOUS ARRAYS

Following in the spirit of research, we look for generalizations of our results.

The array structure characteristic polynomial for an array is built up from its dimensional lengths. These are the entries in its shape vector. What if we considered arrays whose dimensions are measured by a continuous variable? Its lengths would be the integrals of the variable indexing its dimensions. Consider an x by y array.



Its shape vector is: $\{L_x, L_y\}$ where

$$L_x = \int dx \quad \& \quad L_y = \int dy$$

The integrals are taken over an interval of some length so have upper and lower values. The continuous array's characteristic polynomial is:

$$\begin{aligned} T_{L_x, L_y}^2 &= (L_x T_0^0 + T_0^1) \otimes (L_y T_0^0 + T_0^1) \\ &= (L_x L_y) T_0^0 + (L_x + L_y) T_0^1 + T_0^2 \end{aligned}$$

The first term is the count of all its cells, its area. The second term is the count of all of its vectors, horizontal and vertical. Clearly, for any fixed value of $x = X$, we get a horizontal vector, and similarly for $y = Y$ fixed. Also similarly to the discrete case, (X , Y) indexes a cell (a point) in the area. This generalizes to any finite dimensional continuous array. By the way, including the mapping to a co-domain from the continuous array, with the array, is called a "multi-variable function".

8. NEGATIVE LENGTHS

Why stop there? We can investigate negative lengths. We let the elements of the shape vector be continuous variables above. Now we let them be negative values, i.e., negative "lengths".

$$T_{2,-3}^2 = (2T_0^0 + T_0^1) \otimes (-3T_0^0 + T_0^1) = -6T_0^0 - 1T_0^1 + T_0^2$$

The point of this paper is to show how following a trail of development in research can lead to new ideas, so we don't stop here to work out the meaning of negative lengths.

9. COMPLEX LENGTHS

We ask what is the meaning of $T_{a+ib}^1 = (a + ib) T_0^0 + T_0^1$? The answer comes from considering catenation again.

$$\{a\} + \{ib\} = \{a+ib\}, \text{ all 1-D arrays.}$$

$$\begin{aligned}
 T_a^1 &= (a)T_0^0 + T_0^1 \\
 T_{ib}^1 &= (ib)T_0^0 + T_0^1 \\
 (T_a^1), (T_{ib}^1) &= [(a)T_0^0 + T_0^1] + [(ib)T_0^0 + T_0^1] - T_0^1 \\
 (T_a^1), (T_{ib}^1) &= [(a) + (ib)]T_0^0 + T_0^1 = (a + ib)T_0^0 + T_0^1
 \end{aligned}$$

Like complex numbers, complex length arrays combine by computing the real parts and the imaginary parts separately.

$$(T_a^1), (T_{ib}^1) = T_{a+ib}^1 \Rightarrow$$

10. FRACTIONAL DIMENSIONS

Let us now turn to the question of fractional dimensions. We approach this by asking "what is the square root of a 1-vector?"

$$\begin{aligned}
 &(1T_0^0 + T_0^1)^{\frac{1}{2}} \\
 &(1T_0^0 + T_0^1)^{\frac{1}{2}} \otimes (1T_0^0 + T_0^1)^{\frac{1}{2}} = (1T_0^0 + T_0^1)^1
 \end{aligned}$$

From (Knopp 1948 p426 # 247), (Knuth 1997 pp. 53 and following), and (Knuth 1998 pp. 418, 525 and following) we can formally write for the general fraction a:

$$\begin{aligned}
 (1T_0^0 + T_0^1)^a &= \sum_{k=0}^{k=\infty} \binom{a}{k} (T_0^1)^k = \sum_{k=0}^{k=\infty} \binom{a}{k} (T_0^k) \\
 &= \sum_{k=0}^{k=\infty} \left[\frac{a(a-1)\dots(a+1-k)}{k(k-1)\dots(1)} \right] (T_0^k)
 \end{aligned}$$

Infinite dimensionality!

Formally again, this series converges for the variable $T_0^1 \approx x \leq 1$. I. e., the sum of the coefficients converges. Fortunately the terms in brackets alternate in sign so that when we multiply two of these series together for a = 1/2, the terms greater than k=1 in the result all cancel out to zero, leaving only the k=0 and k=1 term, as they should. This result seems to say that fractional dimensional arrays actually exist in an infinite dimensional space. Since fractals often are of fractional dimension, one wonders if there is a connection (??). Let this be a conjecture that it is so.

11. GENERAL FUNCTION OF AN ARRAY

The development for the square root suggests the following conjecture. Let P(A) be the structure polynomial of array A. Let

F(A) be a function of A (such as square root A). Then we conjecture

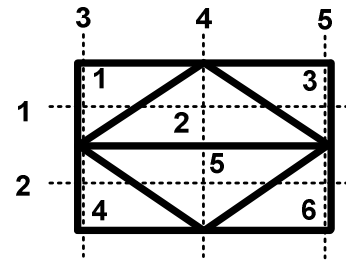
$$P(F(A)) = F(P(A))$$

That is, the polynomial of the function of the array is the function of the polynomial of A. See (Knuth 1998 pp. 526 following) for the conditions on F() that allow this. They involve taking derivatives of F().

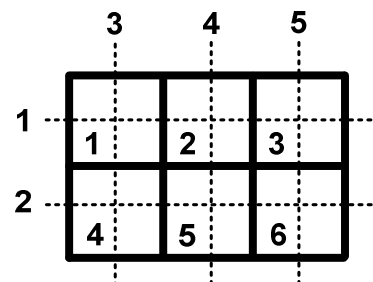
Suffice it to say here that this too works out.

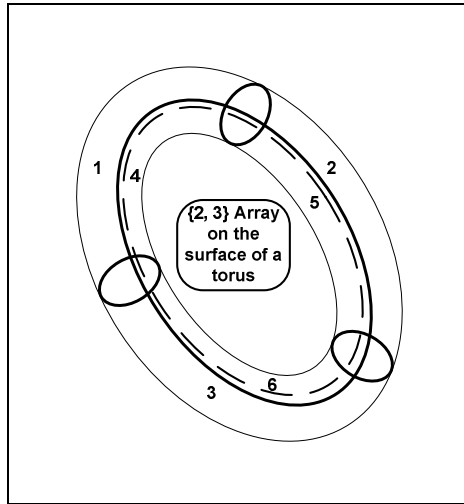
12. QUESTION: WHAT IS THE STRUCTURE OF AN ARRAY? TOPOLOGICAL INVARIANCE

We talk of mapping the cells of an array to a co-domain. The co-domain element is the "contents" of the cell. Everything here has been about the structure without paying attention to the contents of the arrays. There is no a priori shape to a cell other than its associated dimensionality. A 2-D cell could be pictured as a triangle.



There is no a priori reason to consider a two dimensional array to lie in a plane. By identifying the top and bottom edges of the egg crate model, and the left and right edges, we could lay the array on a torus, and still have the same structure polynomial and the same 2-D array. We map the {2, 3} from above onto a torus.





13. QUESTION: IS THERE A COSMOLOGICAL MEANING OF ANTI-GLUE?

If we bring together two continuous 4-dimensional arrays (spaces), they generate a 3-dimensional catenation interface in four dimensions. This is reminiscent of Steinhardt and Turok's alternative to the big bang model (Steinhardt, Paul J. and Turok, Neil. 2007). Since there is nothing in an array specification that determines that array boundaries (whatever THOSE are) are straight lines (see section above), there is no a priori reason not to allow the spaces to interact on a very local (quantum) level. The question then arises as to whether the catenation-created space is related to the creation of matter and energy (vacuum energy). These questions are now outside the normal range of computing data structures.

14. SUMMARY

By making two observations and asking what they mean; by working out the details; by following subsequent questions and working out their details; we come up with an entire set of new ideas and even more interesting new questions and conjectures.

This is the essence of research.

For those practical folks, for whom new truth is not, in itself, enough, let me again point out that slicing and dicing multi-dimensional arrays is the essence of complex parallel

machine architectures and OLAP data "cubes".

Future research directions are indicated in the diagram in the appendix.

16. REFERENCES

- Falkoff, Adin and Iverson, Ken. (June 1970). A private APL tutorial communication.
- Frank, Ronald I. (May 2002) "Regular Array Expansions in Null Arrays with Applications," Dissertation Pace University. ProQuest Digital Dissertations (AAT 3064835)
- Frank, Ronald I. (2003). A Generating Function that Counts the Combinatorial Full-Span Sub Array Structure of a Regular Array. *Proceedings of the ACM SIGAPL Federated Computer Research Conference* (October 2003). ISBN 1-58113-668-4
- Frank, Ronald I. (2003). A New 'Dual-View' Diagram of Array Structure, *Proceedings of the ACM APL (SIGAPL) Federated Computer Research Conference* (October 2003). ISBN 1-58113-668-4.
- Frank, Ronald I. (2003). An Algorithm to Compute All Full-span Sub Arrays of a Regular Array. *Proceedings of the ACM APL (SIGAPL) Federated Computer Research Conference* (October 2003). ISBN 1-58113-668-4.
- Knopp, K. (1948). "Theory and Application of Infinite Series". Hafner NY, NY. [Pre ISBN].
- Knuth, Donald E. (1997). "The Art of Computer Programming" Vol. 1. 3rd Ed. Addison Wesley Reading, MA. ISBN: 0-201-89683-4.
- Knuth, Donald E. (1998). "The Art of Computer Programming" Vol. 2. 3rd Ed. Addison Wesley Reading, MA. ISBN: 0-201-89684-2.
- Steinhardt, Paul J. and Turok, Neil. (2007). "The Endless Universe (Beyond The Big Bang)". Doubleday (Random House) NY, NY. ISBN 978-0-385-50964-0.

APPENDIX A. Research "Tree" with Future Questions to Answer.

