Vector fusion Basics **January 26, 2005**

Vectors and vector-sums

There is a wish to be able to discover, see, understand, and use the relationships that exist in data of many dimensions. Most people “know” that one cannot see data of many dimensions. Part of that belief derives from early training in Cartesian coordinates where 3 orthogonal coordinates are used to describe space in our physical universe. But dimensions are not necessarily only physical coordinates; many different observations, parameters or kinds of measures, or different factors in experiments can be called dimensions. Each column of data in a spreadsheet is called a dimension here, regardless of whatever is being represented in that column of data values. Then, being able to visualize multi-dimensional data first in one complete picture, and then in several pictures with differing perspectives, is helpful because seeing from many angles is an important step in understanding.

What one learns from data without knowing what the individual columns of data represent differs from what one learns using the meaning of each column of data. Logical analyses exploiting combinations of each column’s meanings produce semantic understanding. Semantic understanding is not what vector-fusion is about.

Vector-fusion is a new technique that facilitates the ability to see functional relationships in the multi-dimensional data itself. If the relationships among the several columns of data happened to fit a geometric relationship, such as an n dimensional sphere, one would see an n dimensional sphere by looking at the loci generated with vector-fusion. Since most people do not know what n dimensional geometries look like, especially when encountered in multidimensional data, basic studies of multidimensional geometry have been published [2]. Simpler functional relationships are recognizable by their geometric shape in 2 dimensions. Linear relationships appear as straight lines, trigonometric relationships appear as circles or parabolas or hyperbolas. Cycloidic and more complicated relationships all appear as those geometric shapes. See *Learning from Data* [1] for examples of geometric relationships that arise in real data from oil-well-logs, medical data, and biological data.

Vector-fusion displays the composite functional relationships that exist in multi-dimensional data. That display of the composite relationship is in one complete image for all dimensions. Relationships existing in subsets of dimensions of data can also be discovered by vector-fusing subsets of dimensions. The functional relationships in the data are the relationships that exist relating each dimension one to another, regardless of whether or not those relationships were planned and programmed. Where noise or statistical uncertainty clouds the relationships, curve-fitting techniques can be used to approximate visually obvious loci within the clouds of individual points. Each point in such clouds represents an entire row with the contributions of all of its n dimensional data values.

To accomplish this result, vector-fusion first *vectorizes* and then successively *vector-sums* each subset or row of values of arbitrarily many dimensions of data into one vector-resultant-point. Data typically is presented as scalar values in an array of m rows and n columns. A column is called a dimension. A *vector* is a value or magnitude R at an angle θ as illustrated in Figure 1. *Vectorizing* assigns a different angle to each dimension (ie a different angle is assigned to each of the n columns in the array of data). Thus *vectorized* data is the scalar value in each cell represented as a magnitude at the assigned (phase) angle for that cell’s column.

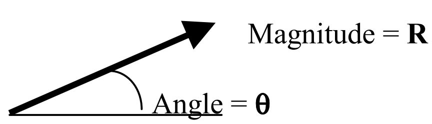


Figure 1. A vector **R** at angle **θ**

It is important to assign a different angle **θ** to each dimension. The properties of each column of data are then the properties associated with all values *at that phase angle*. If for example, moisture content is the semantic meaning of a column, any value of data at the phase angle of that column represents moisture-content. Vector-sums add vectors at their assigned phase angles. Scatter-plots of vector-sums plot data from each column against data values with the same meaning.

Working hypothesis (1): *Mathematical properties associated with each column's data, such as the values of first or second derivatives at each vectorized sample point, are maintained as a property of each dimension in vector-sums of those vectorized dimensions.* In vector-sums, the phase angle for each column of vectorized data is unique to each column of data. See Appendix 3 in *Learning from Data* [1] for more examples demonstrating the association of properties with vectorized dimensions.

An important consequence of Working hypothesis (1) is that the units measuring each dimension of data are irrelevant in the vector-sums. By adding each dimension at its own unique phase angle, one can add (vector-sum) different dimensions one to another without regard to measurement units. Consistent column sequence is important because vector-sums and vector-comparisons (scatter-plots) occur at the phase angles assigned in the original column sequence. Consistent row-sequence is also important to maintain among the vectors.

Vectorizing the data of a column in an array means first to name the values in column 1 as values of vector **V1**. In Figure 2, two vectors **V1** & **V2** are shown with the third resultant vector being the vector-sum of the two vectors. Next, assign a specific angle in display space to each vector, **V1** & **V2**. Thus each row value in column 1 has the value shown in color in **V1** in Figure 2. In this example, Column 1 has 4 values, one value per row, and each row's values for column 1 are given an illustrative color, blue, orange, green, & black. All 4 component vector values of **V1** are pointed in the same direction, that is each row value of **V1** is assigned the same phase angle. The sequence in which the colors are ordered along vector **V1** however is also important.

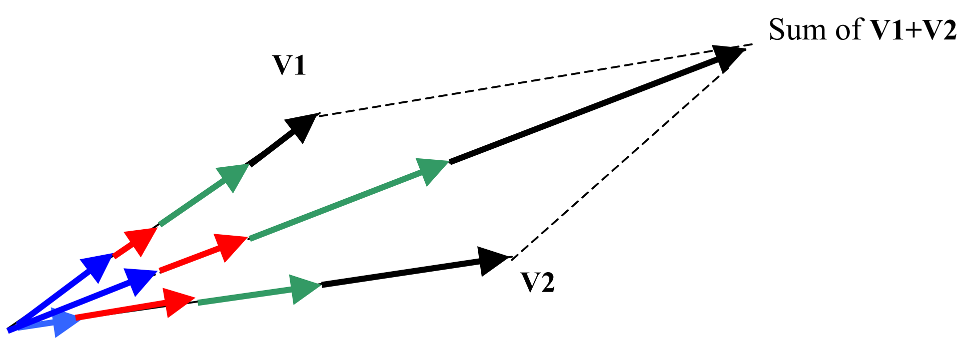


Figure 2 – Vector-sum of 2 vectors

The second column of values in the data array is named vector **V2**. In Figure 2, **V2** also has 4 rows of values, the same color for values in each row of data in **V1** is used for each row in **V2**. The sequence in which the colors in **V2** occur must be the same sequence as the colors appear in V1. The vector which is the vector-sum of **(V1 + V2)** is shown in Figure 2 with the incremental vector-sums for each row colored with the same colors as in **V1** and **V2**.

Each vector **V1** and **V2** has its own different and specific (phase) angle in Figure 2, and the vector sum of **(V1+V2)** has a (phase) angle determined by the values of its component vectors **V1** & **V2**.. The important ideas are the concept of *values at angles*, the concept of *values added together at respective angles*, and the concept of *the sequence* in which the values occur along each vector **V1** & **V2**.

The concept of *phase* angle contains the idea that angles in the display space of Figure 2 are a special kind of angle; each phase angle can be associated with a different kind or dimension of vector. Most importantly, when vector-summing two vectors, the rows (colors) of each vector must occur in the same sequence. Thus when one does a scatter-plot of the (x,y) components of each color as it occurs in the vector-sum, the same colors are compared or plotted against each other. Controlling this (row) sequence of colors is key to understanding why the scatter-plot of the (x,y) components of the vector-sum reveals the functional relationship between or among the values in each of the m rows of different colors in each vector **V1** & **V2**.

As shown in Figure 3, the third column of values from the original array of data is called **V3**. **V3** is then added to the vector-sum of the first two vectors (**V1** + **V2).**

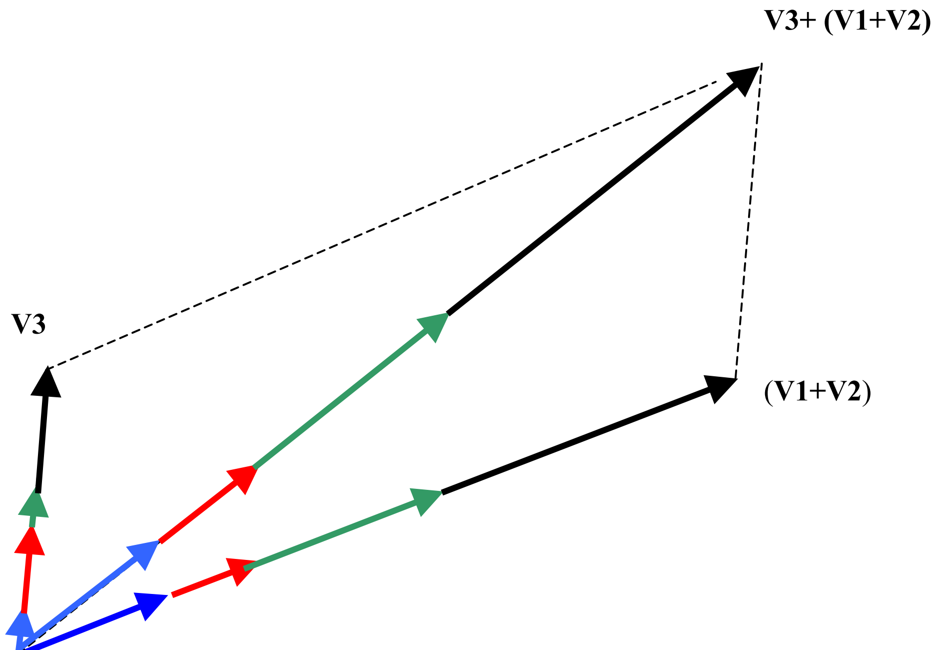


Figure 3. Vector-sum of **V3 + (V1+V2)**

For a big array of data with n columns of values, each column can be assigned its own vector name **Vn**. **Vn** will have a specific, unique, and different phase angle **θn** assigned to each data value in column n. A common row (color) sequence among each and all n columns is important. The row sequence is illustrated in Figures 2 & 3 by the color sequence appearing in each vector **V1** & **V2**. The incremental vector-summing by pairs of vectors as was done in Figure 3 following Figure 2 is continued for all n vectors. The vector-sum for successive pairs of vectors is a final vector sum $\mathbf{Vn}^{+} = [\ \mathbf{Vn} + (\ \mathbf{Vn\text{-}1} + \Sigma_{\iota}\ \mathbf{V}_{i} \text{ for } 1<i< n\text{-}1)\ ]$. $\mathbf{Vn}^{+}$ is computed from all n vectors, one vector representing each of the n columns in the dataset, with each vector being added to the sum of the preceding vectors.

Functional Relationships revealed by Scatter-plots

An important relationship is discovered by plotting (the incremental **x**-component-values) in the vector-sum (**V1+V2**) vs (the incremental **y**-component-values) for each row of data in the vector-sum. This kind of plot is called a scatter-plot. A scatter-plot of a vector with several rows of values plots each row's contribution in the x-coordinate direction of display space vs each row's contribution in the y-coordinate direction. The scatter-plot of the vector which is the vector-sum of (**V1+V2**) is the plot of (**V1x+V2x**) vs. (**V1y+V2y**) for each row of values**.** This scatter-plot reveals the functional relationship that exists between the (x,y) components of each row (and in their color sequence).

In Figure 2 and Figure 3, the magnitude (length or value) of each color is randomly chosen. Figure 4 has the values of data for Figures 2 &3; the phase angle assignments for Columns 1,2,3; and the vector sum values for (**V1+V2**). Figure 4 includes the scatter-plot for the data of Figure 2.

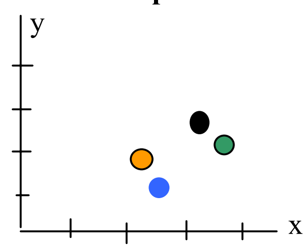
Data values of Fig.2&3:

Col.1	Col.2	Col.3
2	1	1
1	2	2
2.5	2	1
1.5	3	3

Phase angle assignments:

Col.1	Col.2	Col.3
10^{o}	50^{o}	85^{o}

Scatter-plot of V1+V2:



The vector-sum (**V1+V2)** is shown diagrammatically in Figure 2. The x and y coordinate values for **V1** and **V2** are found using the cosine and sine values for 10^{o}, 50^{o} and 85^{o} :

V1|**x** +V2|**x**	V1|**y** +V2|**y**	row color in Figure 2 and scatter-plot above:
2.6	1.01	blue
2.26	1.71	orange
3.73	1.96	green
3.39	2.54	black

The scatter-plot of these 4 row (x,y) values is the colored plot just above.

Figure 4

With the random relationship between the values of data for V1,V2,V3 in Figures 2,3,4 the scatter-plot for (**V1+V2+V3**) would again be a plot of 4 points in (x,y) display or vector space, a "cloud" of 4 randomly spaced points. The "cloud" in the scatter-plot of Figure 4 is the visually displayed functional relationship among the 2 dimensions represented by **V1** & **V2**.

The scatter-plot of the (x,y) vector-components of the row values in the vector-sum reveals the precise functional relationship that exists between the values of **V1** and **V2**. If there were a linear relationship between the row-values in each column (or dimension) of data, the scatter-plot of the vector-sums (by row) would be a straight line.

Figure 5 shows a sinusoidal vector **x** with 24 values of data and a cosinusoidal vector y as the second vector **y**. The scatter-plot of the vector-sum of these two sinusoids plotted orthogonally (at phase angle = Pi/2) is the circular arc shown in Figure 6.

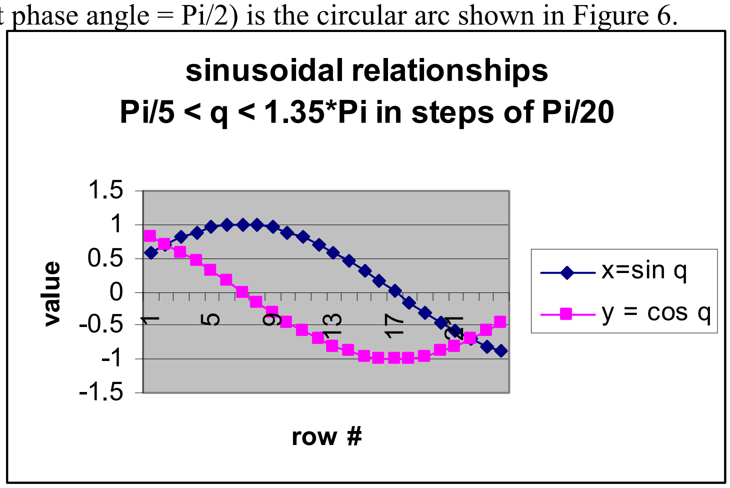


Figure 5. Sinusoidal relationships in 2 dimensions of data

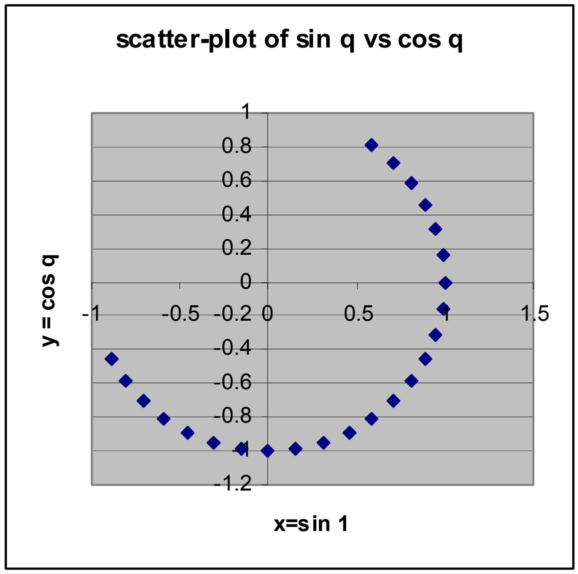


Figure 6. Scatter-plot of sin vs cos at phase angle=Pi/2

The functional relationship between the two vectors of 24 points of data in Figure 5 is the circular relationship shown in Figure 6 when the two vectors **x** and **y** are plotted orthogonally.

When the same two vectors **x** and **y** are plotted at a phase angle = Pi/6 = 30^{o} with respect to each other, the functional relationship is the ellipse displayed in Figure 7.

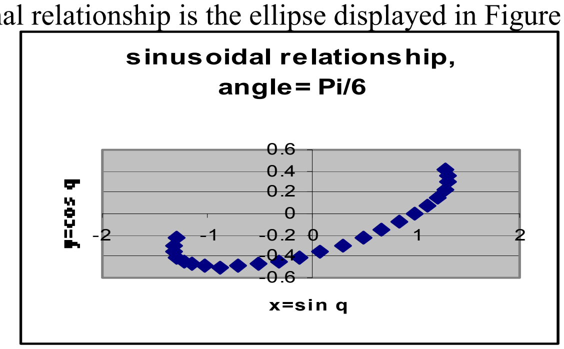


Figure 7. Functional relationship displayed at phase angle = Pi/6

At other acute phase angles between the two sinusoidal vectors **x** & **y**, the functional relationships displayed in the scatter-plots of the two vectors are ellipses with varying eccentricities.

Figure 8 is an example 12 dimensional dataset containing 6 different cardioids each in 2 dimensions with 200 rows of values for each cardioid. The six cardioids have diameters that vary by 20:1 and different orientations and offsets.

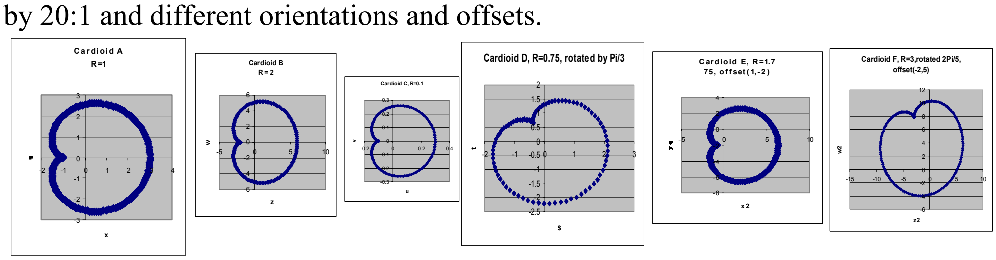


Figure 8. 6 Different cardioids

Figure 9 is the scatter-plot of the SBP vector-sum of the 12 dimensional data with one vector-resultant point for each of the 200 rows of data. The SBP vector-fused Functional Relationship for all 6 different cardioids is one composite cardioid.

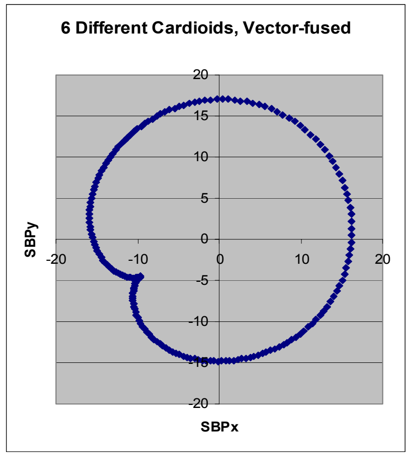
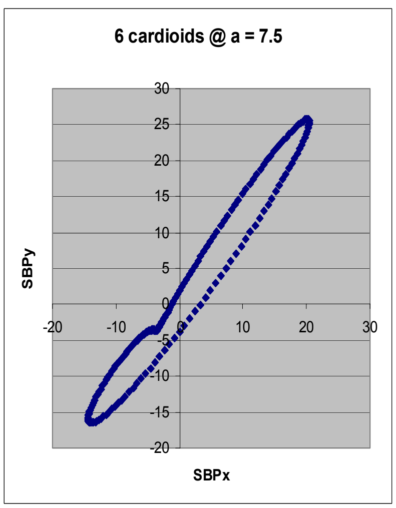


Figure 9. Vector-fused SBP Functional Relationship of all 6 different cardioids.



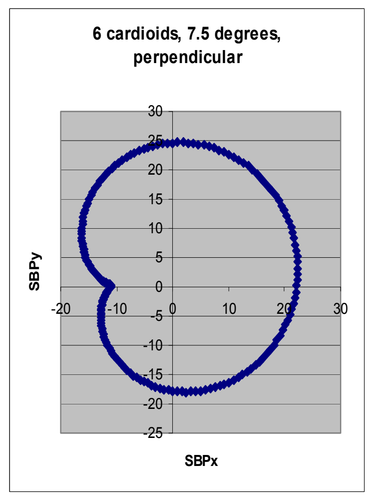


Figure 10A. Phase-angles all equal @ 7.5° Figure 10B. Phase-angles alternate perpendicular

Figure 10A is the same 6 cardioids summed with the vector-fused scatter-plot (SBPx, SBPy) but with all vectorized phase angles at equal increments of 7.5°. In contrast, as for the data behind Figure 9, SBP vectorizes each odd numbered Dimension with an incremental phase angle = 15° AND with even (or alternating) Dimensions being placed at 90° or perpendicular to the intervening odd Dimensions. Figure 10B is the scatter-plot using variously different phase angle assignments (6°, 12°, and 24°) but with alternate (even numbered Dimensions) phase-angle assignments being 90° just as is done normally in SBP.

The significance of Figure 10B is first, that geometry is preserved when vectorizing each dimension of data by assigning phase-angles of 90° to even dimensions and incremental acute angles (of any value). Second, the same shaped geometry is displayed regardless of the phase angle assignment as long as alternate Dimensions are perpendicular to each other. Any phase angle can be used if even Dimensions are 90° and odd Dimensions are arbitrary as long as the total sum for all n Dimensions equals 180°.

If several different 2D geometries are present in the multidimensional data, the vector-fused resultant is the vector-sum composite of those geometries.

Most importantly, the scatter-plot of the final vector-sum $\mathbf{Vn}^+$ reveals the shape of the composite functional relationship that exists among all n vectors of the multi-dimensional dataset. The shape displayed by the scatter-plot remains the same for similar geometries within and among the n dimensions of data as long as:

1. The curvature of each internal geometry is the same,
2. Alternate (even) dimensions are perpendicular to the preceding (odd) dimension. The preceding dimensions can be assigned any acute angle subject to the total sum not exceeding 180°. If all odd dimensions are assigned acute phase angles, the structure revealed by the vector-fused resultants is the structure of the family of geometries. For example, ellipses are the family of shapes for circles; circular Functional Relationships are produced as the Functional Relationships when alternate dimensions are perpendicular to each other; otherwise, ellipses are the Functional Relationship structures produced by the vector-summing process.

This property of the final vector-sum $\mathbf{Vn}^+$, for all n vectors, is best understood by considering the sums incrementally; each successive pair of vectors added together can be scatter-plotted. The two incremental component vectors ($\mathbf{Vn\text{-}1}_x$, and $\mathbf{Vn\text{-}1}_y$) define $\mathbf{Vn\text{-}1}^+$ in the (x,y) coordinates of display space. Each incremental vector-sum is in particular the vector-sum of all of its m (colored) (x,y) component vectors, dimension-by-dimensional sum, in color sequence by row. Scatter-plots of each incremental vector-sum pair reveal the functional relationship that exists for that pair of vectors.

Working hypothesis (2): *With varying phase angle assignments to each vector, similar shapes of the functional relationship are displayed in scatter-plots for geometric shapes having equal curvatures. Curvature, k, is defined by the derivatives or tangents to the loci represented by each vector.* [3]

Thus, a scatter-plot of the (x,y) components in the final vector-sum $\mathbf{Vn}^+$ reveals, for all m points in its scatter-plot, the composite structure of the functional relationships that exist in the data of all n dimensions and for all m rows of that data.

Hypothesis 2 summarizing the results of Figures 9 and 10, and the reasonings above, says that the single-point vector-sum resultant of each row of n Dimensional data represents the properties and contributions of each of the n Dimensions. Thus with row synchronized (phase-angle synchronized) vectorized multidimensional data, the Functional Relationship displayed by scatter-plots of the m points' component resultants in the x direction, SBPx, and in the y direction, SBPy, is the shape of the composite geometric structure in the entire dataset regardless of incremental phase angle assignment.

Irrelevance of unit measurements of each Dimension

What **is not** important in the diagrams of vectors and vector-sums is the units by which values in each column are measured. As long as blue values are added to blue values, orange values added to orange values, etc., and the vector-sums are always at the respective phase angles of their own two vectors, the vector-sum captures whatever units are represented in each of the colors blue, orange, green, and black. Same-values (colors) are always added to each other in the vector-sum. In the vector-sum, each colored value is always accumulated at the phase-angle of its respective dimension (ie column in the data array) and in the same color (row) sequence for each vector.

What **is** important in these diagrams of vectors is that the vector-sum represents the colored values in each row that are accumulated together, and accumulated *at the phase angle* of each respective column (**V1** and **V2**). Each vector must have its rows in the same (color) sequence. The vector-sum accumulates the total value and phase angle for each pair of vectors.

Conclusions

The vector-sum uses, captures, and represents the contributions (and properties) of each vectorized dimension because each vector (its values each at its unique phase angle) represents the measures, units, and properties of each dimension. This property or ability of each vector arises because each vector's contributions are always accumulated in row sequence and functionally compared in the scatter-plots at each vector's unique and specific phase angle.

The scatter-plot of the m row values of the x and of the y components in the final vector-sum of the vectors of all n dimensions of data reveals the shape of the locus of those m points. The shape of the locus is the composite functional relationship that exists for all m rows of the n-dimensional data.

As long as alternate Dimensions are vectorized at phase angles perpendicular to the preceding Dimension, the shape of the Functional Relationship appears independent of the value of the phase angle assignment. This independence has been demonstrated for multiple linear, circular, cardioid, and catenary geometries within the multidimensional data.

A mathematical proof of this inference about independence has not yet been done even though many numerical examples have been tested which lead to the conclusions above.

Further examples of functional relationships found in real and in mathematically generated data are given in *Learning from Data* [1].

R.R. Johnson
nDV

References

[1]. *Learning from Data* in www.n-dv.com
[2]. Johnson, R.R., Millar, R., Ben, I., "Visual Identification of Structure of Data," *SPIE Conference on Visualization* 1 (Jan. 2003).
[3]. Oprea, J., *Differential Geometry and its Applications*, Prentice Hall, 1997, page 19