Evaluation of Star Coordinate Boundaries

Yan Chao Wang
Rolls-Royce@NTU Corporate Lab,
Nanyang Technological University
Singapore
ywang054@e.ntu.edu.sg

Feng Lin
Nanyang Technological University
Singapore
ASFLIN@ntu.edu.sg

Hock Soon Seah
Nanyang Technological University
Singapore
ASHSSEAH@ntu.edu.sg

ABSTRACT

In the visual analysis of high-dimensional data points, star coordinate plot maps the data points into an observable space by conducting a linear combination of radial distributed axes vectors. However, when plotting additional data points which are quite common in an engine performance and simulation model evaluation, the star coordinate is not effective due to undetermined plotting boundary. In response, we propose a geometric algorithm to determine the axis-aligned bounding box, minimum bounding box, and bounding polygon of the star coordinate with an evenly distributed configuration and an arbitrary distributed configuration respectively. These boundaries enable a continuous point rendering that adapts to the screen space boundary. Moreover, the determination of the boundaries provides deeper understandings on the star coordinate plots. Firstly, we amend the expression of the star coordinate by considering the boundaries such that the plotting can be generated reasonably. Secondly, the boundaries provide an objective and quantitative transformation to the widely accepted RadViz measures, in order to exploit the advantages of both. Finally, property analysis, such as the clumping effect of the star coordinate plot, is conducted based on the calculated boundaries.

CCS CONCEPTS

• **Human-centered computing → Visualization theory, concepts and paradigms**; *Visualization techniques*;

KEYWORDS

Star Coordinate, Subsequent Data, Visual Analytics, Bounding box, Bounding polygon

ACM Reference Format:
Yan Chao Wang, Feng Lin, and Hock Soon Seah. 2018. Evaluation of Star Coordinate Boundaries. In *CGI 2018: Computer Graphics International 2018, June 11–14, 2018, Bintan Island, Indonesia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3208159.3208181

1 INTRODUCTION

Extensive studies on visualization of the properties of high-dimensional and large-scale datasets greatly benefit exploratory data analysis. In our case of engine performance and simulation model evaluation, effective visualization methods can provide engineers the ability to view, explore, filtrate, compare and comprehend trend and structure of the high-dimensional engineering datasets in an observable space. These visualization methods are initially designed to gain insights from the plotting, and the recent trend is more on indicative analyses with additional data points.

Star coordinate (SC) plot [14] adopts a radial pattern to conduct data analysis. It has emerged as a promising visual mapping approach in recent years. By transforming data point with a combination of radial distributed axes vectors, SC linearly maps data points with a high computational performance. These characters indicate great potential of SC to provide insights on high-dimensional data visualization. Nevertheless, such potential is yet to be translated to the capability in real applications due to insufficient works. For instance, the cluster representation should include outlier detection and attribute estimation, which is still imperfect in SC. A technical challenge is that data points which are far away from each other in high-dimensional space may be mapped to the nearby points in low-dimensional plotting. A more challenging problem is that the range of SC for all the possible projected data points is not determined. This range not only enables the plot of additional data points based on current configuration but also supports the analysis of the geometrical properties of SC. These lead to a core computational problem of SC boundary evaluation. To the best of our knowledge, no research work on the boundary of SC plots has been published.

In this paper, we propose a generic solution to the boundary evaluation of SC plot, which has standard and arbitrary configurations, by introducing optimal configuration for special datasets. To start with, we propose a geometric algorithm to determine the boundary of SC with standard configuration which means all the axes are evenly distributed. The boundary includes the axis-aligned bounding box and bounding polygon. The detailed algorithm on how to calculate the bounding polygon, is illustrated by an example of a 7-dimensional dataset. Then, the algorithms to calculate the axis-aligned bounding box, the bounding polygon, and the minimum bounding box of SC with arbitrary configurations, as well as geometric character analysis, are described. These boundaries enable the plot of subsequent data points based on the current configuration while not having to dynamically adjust the display window. Moreover, the determination of the boundary benefits the understanding of SC plot in several ways. Firstly, based on the determined boundaries, we amend the expression of SC. By using the new expression, not only subsequent data points can be plotted appropriately, but also the window size can be customized with arbitrary configurations. Secondly, the SC plot can be associated with the well accepted visualization method RadViz [10]. The boundaries provide a fair comparison between RadViz and the SC by fixing the

CGI 2018, June 11–14, 2018, Bintan Island, Indonesia

ACM ISBN 978-1-4503-6401-0/18/06…$15.00
https://doi.org/10.1145/3208159.3208181

plotting size, are helpful to rebuild the relationship between these two popular plotting methods and make use of the advantages of them. Finally, the boundaries enable the property analysis of SC plot as the dimensionality increases. One example is the analysis on clumping effect which is often known as the effect that forces data points towards the center of the plotting during dimension reduction. As the range of SC changes along the dimensionality, the determination of boundary makes the clumping effect analysis clear.

The main contributions of this paper are:

- we are first to calculate the boundaries of SC plot with standard and arbitrary configurations as well as some geometric properties of the boundaries;
- we amend the expression of SC with arbitrary configuration for subsequent plotting and customized window size;
- we reformulate the relationship between RadViz and standard SC by considering the SC boundary;
- we analyze the clumping effect of SC with standard configuration.

The rest of this paper is organized as follows: the related work is reviewed in Sect. 2. Sect. 3 describes the boundaries of SC with standard configuration as well as related analysis. Sect. 4 presents the boundaries of SC with arbitrary configuration. Applications to utilize the findings based on the boundaries are given in Sect. 5. Finally, discussion and conclusion are presented in Sect. 6.

2 RELATED WORK

The radial visualization method, SC [14], linearly maps high-dimensional data points into low-dimensional space by taking all dimensionalities into consideration. The mapping bases on the standard assumption that the values of each data point should be normalized into the interval of $[0, 1]$ first. This normalization step for an n-dimensional dataset $D = (d_1, \ldots, d_j, \ldots, d_m)$ which contains m data points transfers data point $d_{i,j}$ to $d'_{i,j}$ for $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$. Each dimensionality axis is represented by using a low-dimensional vector $\overrightarrow{v}$. Then the projected point $p_j^{SC}(x_j, y_j)$ in 2D SC is simply the linear combination of these radial distributed vectors, where the linear coefficients are the attributes of each data point. Formally, for SC:

$$p_j^{SC}(x_j, y_j) = \left(\sum_{i=1}^{n} (d'_{i,j} \cdot L_i \cdot \cos\theta_i), \sum_{i=1}^{n} (d'_{i,j} \cdot L_i \cdot \sin\theta_i) \right), \quad (1)$$

where (x_j, y_j) is the projected data point p_j^{SC} in two-dimensional space and (L_i, θ_i) is the magnitude and angle of the i^{th} vector $\overrightarrow{v_i}$. For standard SC, all the radial distributed vectors have unit magnitude and are evenly distributed which means that the angle between successive vectors is the same.

The main disadvantage of standard SC is overlapping projected data points and the phenomenon that different data points in high-dimensional space may be projected close together in SC. To solve this problem, techniques including interaction techniques to control the axes vectors [15], orthogonal projections [17] and heuristics algorithm for dimension ordering [3] are proposed. The kernel of these methods is to find the desired configuration of the radial distributed vectors $\overrightarrow{v_i}$. Hence, SC with arbitrary configuration is more commonly used in practical applications to generate visual representations.

Star Coordinate plotting not only aims to provide an overview of high-dimensional datasets but also often is integrated into the data analysis process with the exploratory purpose including cluster representation [5, 14, 25, 30, 33], outlier detection [23], and data attribute estimation [24]. As SC is a linear projection, a bunch of linear dimensionality reduction methods can be combined with SC to get a better visual interpretation of the clusters in high-dimensional space. Popular methods including principal component analysis (PCA) [21], linear discriminate analysis (LDA) [8, 16], large margin nearest neighbor (LMNN) [32], and neighborhood component analysis (NCA) [9] can be reproduced by choosing proper axes vectors. These have been successfully implemented in order to preserve distance and separate class or clusters in SC [23]. Besides reproduced classification results, optimal sets of axes vectors have been studied to get maximal insight [18] and trial to improve data attribute estimation in SC by implementing data centering, see [24]. Essentially, the kernel is to calculate the magnitudes and angles of radial distributed axes vectors based on certain exploratory purpose. With different configurations of axes vectors, different insight will be obtained and the range of the plotting may also be different. To enhance the capability of visualization, some works also focus on 3D visualization method development [2, 12, 27].

Another radial visualization method often compared with SC is RadViz [10]. RadViz described by using a spring model is well accepted and has broad applications [1, 4, 11, 12, 19, 20, 26, 28, 31]. The main difference between SC and RadViz is that RadViz defines a non-linear mapping. So linear transformations which can be combined with SC are difficult to be implemented in RadViz. The expression of RadViz can be defined as:

$$p_j^{RV}(x_j, y_j) = \left(\frac{\sum_{i=1}^{n} (d'_{i,j} \cdot L_i \cdot \cos\theta_i)}{\sum_{i=1}^{n} d'_{i,j}}, \frac{\sum_{i=1}^{n} (d'_{i,j} \cdot L_i \cdot \sin\theta_i)}{\sum_{i=1}^{n} d'_{i,j}} \right), \quad (2)$$

where $p_j^{RV}(x_j, y_j)$ is the projected data point in two-dimensional RadViz. Then the relationship between SC and RadViz can be expressed as:

$$p_j^{RV}(x_j, y_j) = \frac{1}{\sum_{i=1}^{n} d'_{i,j}} \cdot p_j^{SC}(x_j, y_j). \quad (3)$$

The non-linear factor $1/\sum_{i=1}^{n} d'_{i,j}$ in Eq. 3 brings new properties into RadViz. The theoretical justification for the properties of RadViz lays a foundation for future visualization research [6]. In RadViz, each dimensionality is associated with a point on a unit circle which is called a dimensional anchor. One property is that RadViz maps each high-dimensional data point to a point that is within the convex hull of the dimensional anchors while all dimensional anchors are located on a circle with unit radius [6]. The circumscribed circle of RadViz is fixed for any plotting even the location change of dimensional anchors. It is convenient to give the user an overview of the whole high-dimensional space by using RadViz. Not only for RadViz, other famous high-dimensional data visualization methods including scatter plot matrix (SPLOM), parallel coordinates plotting (PCP) [13], and heatmap [22] are all with fixed range for all the potential projected data points.

However, the range of SC is not determined and keeps increasing as the dimensionality increases. The current solution for this problem when plotting with SC is that a region that can cover all the projected data points is chosen as the boundary of SC. There are a few drawbacks as such. Firstly, this operation leads to the limitation that the user will lose the global view of the whole data space. When analyzing the properties of standard SC, this limitation may lead to unfair conclusions (e.g. the relationship between SC and RadViz (Sect. 5.2) and clumping effect in SC (Sect. 5.3)). Secondly, subsequent data points may locate outside the selected region. The current solution may lead to the continuous change of the view pattern when plotting subsequent data points which are inefficient and may generate confusion.

3 BOUNDARY OF SC WITH STANDARD CONFIGURATION

The boundaries of SC with standard configuration, including axis-aligned bounding box and bounding polygon, are studied in this section. The SC with standard configuration is defined, such that the magnitude of radially even distributed axes vector is one. Thus, following Eq. 1, in standard SC, $L_i = 1$ for $\forall i$. The value of each angle θ_i is

$$\theta_i = \frac{2\pi}{n} \cdot (i-1) + \theta_1 \qquad i \in \{1, \dots, n\}, \tag{4}$$

where θ_1 refers to the angle of the first axis vector. The range of θ_i is $[0, 2\pi)$. As *sin* and *cos* functions are periodic, θ_i other values can be converted into this interval. In this paper, without loss of generality, the assumption that in standard SC the angle of the $\overrightarrow{v_1}$ is equal to zero ($\theta_1 = 0$) is made.

3.1 Axis-aligned Bounding Box

The bounding box that aligns with the axes of the Cartesian coordinate system is known as the axis-aligned bounding box (AABB). AABB as the simplest bounding volume is commonly used to contain more complex objects. We study the AABB of standard SC first.

Proposition 3.1. *Let* $[X_{\max}, X_{\min}, Y_{\max}, Y_{\min}]$ *be the right, left, top, and bottom side of the AABB which contains the whole SC plotting. In standard SC when plotting an n-dimensional dataset, the four sides of the AABB coordinates are*

$$\begin{cases} X_{\max} = \sum_{i \in I} \cos\theta_i & with \quad I = \{i \in \mathbb{N} | \cos\theta_i \geq 0\} \\ X_{\min} = \sum_{i \in I} \cos\theta_i & with \quad I = \{i \in \mathbb{N} | \cos\theta_i \leq 0\} \\ Y_{\max} = \sum_{i \in I} \sin\theta_i & with \quad I = \{i \in \mathbb{N} | \sin\theta_i \geq 0\} \\ Y_{\min} = \sum_{i \in I} \sin\theta_i & with \quad I = \{i \in \mathbb{N} | \sin\theta_i \leq 0\} \end{cases}, \tag{5}$$

where $\mathbb{N} = \{1, \cdots, n\}$.

Proof. Here, we take $X_{\max}$ as an example. To get the maximum value in the positive direction, all vectors $\overrightarrow{v_i}$ are projected onto the x-axis. According to Eq. 1, if $\cos\theta_i > 0$, then set the corresponding value $d'_{i,j}$ to one. Otherwise, set the $d'_{i,j}$ to zero. The values of $X_{\min}$, $Y_{\max}$, and $Y_{\min}$ can be obtained in a similar way. □

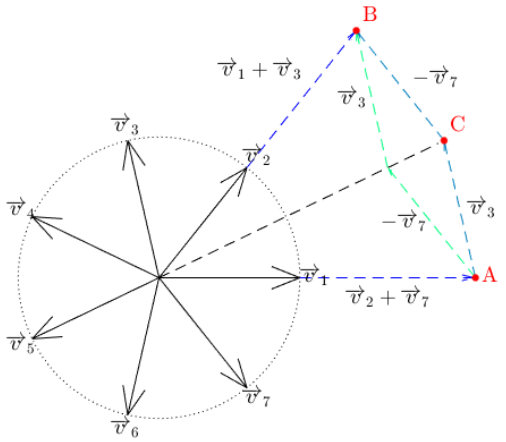


Figure 1: For a 7-dimensional standard SC, the two possible vector orders from point A to point B are illustrated.

Corollary 3.2. *In standard SC, AABB can also be determined as*

$$\begin{cases} X_{\max} = \frac{1}{2}\sum_{i=1}^{n} |\cos\theta_i| \\ X_{\min} = -\frac{1}{2}\sum_{i=1}^{n} |\cos\theta_i| \\ Y_{\max} = \frac{1}{2}\sum_{i=1}^{n} |\sin\theta_i| \\ Y_{\min} = -\frac{1}{2}\sum_{i=1}^{n} |\sin\theta_i| \end{cases}. \tag{6}$$

Proof. As it is an evenly distributed configuration in standard SC, the high-dimensional data point $d'_j = \{1, \dots, 1\}$ will be projected to the center of SC regardless of the dataset dimensionality. Hence, $X_{\max} = -X_{\min}$ and $Y_{\max} = -Y_{\min}$. □

According to the statement in Corollary 3.2, when plotting dataset with higher dimensionality, the area of the SC AABB is significantly larger than that for lower dimensionality dataset. With unit axes vector $\overrightarrow{v}$, for the 4D case, the AABB is $[1, -1, 1, -1]$ while it is $[2, -2, \sqrt{3}, -\sqrt{3}]$ for the 6D case based on Eq. 6.

3.2 Bounding Polygon

The AABB of standard SC is helpful to determine the size of the display window. It encloses the exact plotting region of SC. The more complex bounding polygon that covers all projected data points in standard SC is studied. It is well known that in RadViz all the data points are located in a convex hull formed by all the dimensionality anchors [6]. However, in standard SC, a polygon region that encloses all the possible projected data points has not been studied before. The determination of this bounding polygon region will provide support to the property analysis of SC.

3.2.1 7D Standard SC. Here, our analysis is based on standard SC and the case to plot a 7-dimensional (7D) dataset is taken as an example to illustrate the procedure.

As standard SC is symmetric, instead of finding the whole bounding polygon directly, our analysis starts from the boundary between two successive axes vectors. In Fig. 1, $\overrightarrow{v_1}$ and $\overrightarrow{v_2}$ are chosen to be studied first. In Fig. 1, the farthest point along $\overrightarrow{v_1}$, A, equals to $\overrightarrow{v_1} + (\overrightarrow{v_2} + \overrightarrow{v_7})$. Meanwhile, the farthest point along $\overrightarrow{v_2}$, B, equals to

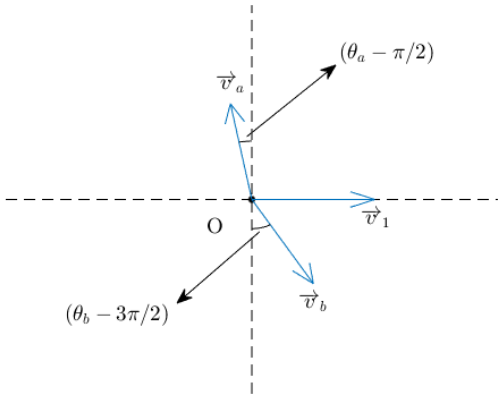


Figure 2: Determine the order of $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$ for $\overrightarrow{v_1}$.

$\overrightarrow{v_2}+(\overrightarrow{v_1}+\overrightarrow{v_3})$. For points A and B, according to Eq. 1 the corresponding high-dimensional data points can be $d'_A = (1,1,0,0,0,0,1)$ and $d'_B = (1,1,1,0,0,0,0)$. Then the polygon region of star coordinate between the direction of $\overrightarrow{v_1}$ and $\overrightarrow{v_2}$ can be considered as the movement from point A to point B. There are two possible solutions for this movement which may lead to a different track. The first solution is $A+\overrightarrow{v_3}-\overrightarrow{v_7}=B$ while the second one is $A-\overrightarrow{v_7}+\overrightarrow{v_3}=B$. These two possible solutions are plotted in Fig. 1 with different colors.

Here, in this case, the first solution is the correct result as $\overrightarrow{v_3}$ plays a more important role at the beginning. This result can be explained more clearly by using the point C in Fig. 1. Point C can be obtained by $\overrightarrow{v_1}+\overrightarrow{v_2}+\overrightarrow{v_3}+\overrightarrow{v_7}$. Hence, $\overrightarrow{v_3}$ connects point A with point C. Similarly, $-\overrightarrow{v_7}$ connects point C with point B. Thus, the bounding polygon between the direction of $\overrightarrow{v_1}$ and $\overrightarrow{v_2}$ is $(\overrightarrow{v_3}, -\overrightarrow{v_7})$.

Using the rotational symmetry property of standard SC, the rest part of the polygon region can be generated easily. In all, the polygon region is formed by vectors $(\overrightarrow{v_3}, -\overrightarrow{v_7}, \overrightarrow{v_4}, -\overrightarrow{v_1}, \overrightarrow{v_5}, -\overrightarrow{v_2}, \overrightarrow{v_6}, -\overrightarrow{v_3}, \overrightarrow{v_7}, -\overrightarrow{v_4}, \overrightarrow{v_1}, -\overrightarrow{v_5}, \overrightarrow{v_2}, -\overrightarrow{v_6})$ which are successively connected.

3.2.2 n-D Standard SC. For the case with an n-dimensional dataset, the bounding polygon can be determined similarly as shown in Fig. 1. In the n-dimensional case, we start from $\overrightarrow{v_1}$ to find the bounding polygon. Let V_i be axes vectors that have positive dot product with $\overrightarrow{v_i}$ and $\overline{V_i}$ is complement of V_i. The $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$ for $\overrightarrow{v_i}$ is defined as $\overrightarrow{v_a} \in (\overline{V_i} \cap V_{i+1})$ and $\overrightarrow{v_b} \in (V_i \cap \overline{V_{i+1}})$. Then, the problem to find the bounding polygon can be divided into three steps: determine the position of the starting point, find the $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$ for each $\overrightarrow{v_i}$, and calculate the order of $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$.

- Determine the position of the starting point. The farthest point along $\overrightarrow{v_1}$ can be calculated as the sum of axes vectors in V_1. In the 7D case in Sect. 3.2.1, $V_1 = \{\overrightarrow{v_1}, \overrightarrow{v_2}, \overrightarrow{v_7}\}$.
- Find the $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$ for each $\overrightarrow{v_i}$. For $i = 1$, according to the definition of $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$, they can be expressed as $\overrightarrow{v_a} \in (\overline{V_1} \cap V_2)$ and $\overrightarrow{v_b} \in (V_1 \cap \overline{V_2})$. In the 7D case in Sect. 3.2.1, $V_2 = \{\overrightarrow{v_1}, \overrightarrow{v_2}, \overrightarrow{v_3}\}$. Then $\overrightarrow{v_a} = \overrightarrow{v_3}$ and $\overrightarrow{v_b} = \overrightarrow{v_7}$. Proposition 3.3 proves that for standard SC $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$ are unique. The determination of $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$ in n-D standard SC is described in Proposition 3.4.
- Calculate the order of $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$. The influence on $\overrightarrow{v_1}$ will be analyzed to determine the order of $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$. Proposition 3.5 gives the criteria used for this determination.

Proposition 3.3. *There is one and only one vector $\overrightarrow{v_i}$ that satisfies the requirement $\overline{V_1} \cap V_2$. Similarly, there is one and only one vector $\overrightarrow{v_i}$ that belongs to the set of $V_1 \cap \overline{V_2}$.*

Proof. This is due to the symmetric configuration of standard SC. □

Proposition 3.4. *Let θ_1 be the angle of axes vector v_1. Then based on θ_1, the $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$ in standard SC can be expressed as:*

$$\theta_a = \begin{cases} \frac{\pi}{2} + \theta_1, & \text{if } n = 4\mathbb{Z}^+, \\ (int\left[\frac{n}{4}\right] + 1) \cdot \frac{2\pi}{n} + \theta_1, & \text{otherwise}, \end{cases}$$

$$\theta_b = \begin{cases} 2\pi - (\frac{n}{4} - 1) \cdot \frac{2\pi}{n} + \theta_1, & \text{if } n = 4\mathbb{Z}^+, \\ 2\pi - (int\left[\frac{n}{4}\right]) \cdot \frac{2\pi}{n} + \theta_1, & \text{otherwise}, \end{cases}$$

where $\mathbb{Z}^+$ refers to a positive integer, $int[x]$ gives the largest integer less than or equal to x.

Proof. Let $\theta_\Delta = \theta_i - \theta_1$, $i = 1, \ldots, n$ and θ_i is defined as in Eq. 4. As i increases, θ_a is the first one that satisfies the condition: $\theta_\Delta = \theta_a - \theta_1 \geq \pi/2$. While for θ_b, it is the first one that satisfies the condition: $\theta_\Delta = \theta_a - \theta_1 > 3\pi/2$. In low-dimensional cases, it is possible that $\theta_b = \theta_1$. □

Proposition 3.5. *The influences of θ_a and θ_b on θ_1 determine the vector order. If*

$$\theta_b - 3\pi/2 > \theta_a - \pi/2,$$

then the vector order of $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$ is $(\overrightarrow{v_a}, -\overrightarrow{v_b}, \cdots)$. Otherwise, the order should be $(-\overrightarrow{v_b}, \overrightarrow{v_a}, \cdots)$.

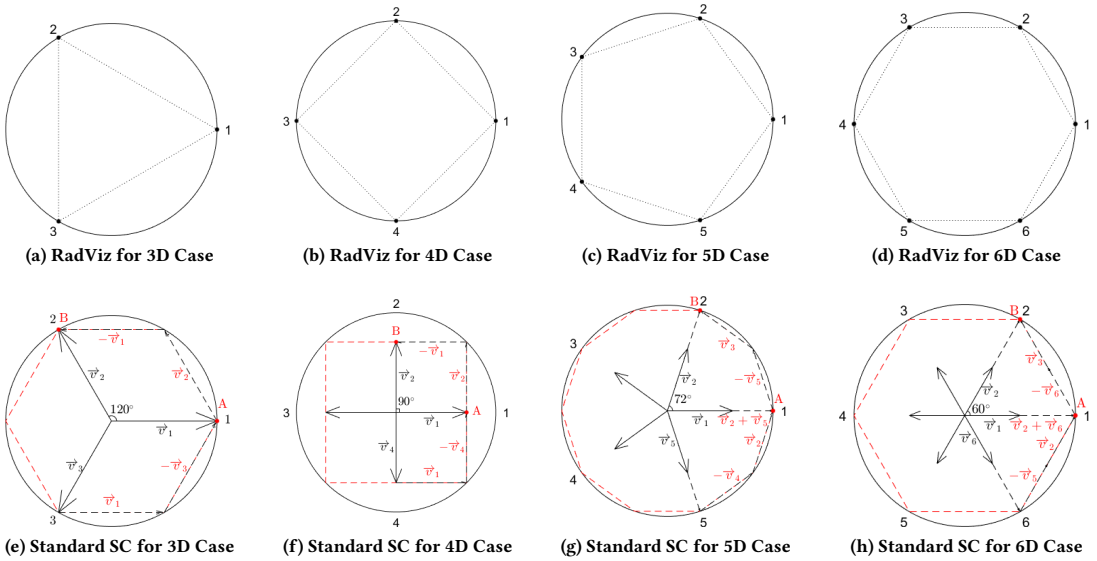
Proof. The included angle between θ_a and $\pi/2$ and the included angle between θ_b and $3\pi/2$ are compared (as shown in Fig. 2). □

Once these three issues are solved, the remaining vectors can be arranged successively, and the bounding polygon obtained. A pseudo-code is given in Algorithm 1.

3.2.3 Geometrical Property. Here, we show that the projected data points in standard SC lie within a convex hull.

Proposition 3.6. *When transforming high-dimensional data point d'_j to $p_j^{SC}(x_j, y_j)$, the bounding polygon of standard SC is a convex polygon.*

Proof. The convexity of a bounding polygon of standard SC can be proved by considering the definition of convexity. In the set of vector V, if for any two vectors $\overrightarrow{v_1}, \overrightarrow{v_2} \in V$ and any $t \in [0,1]$, the vector $t\overrightarrow{v_1} + (1-t)\overrightarrow{v_2}$ is also in V, then it is said to be convex [7]. According to Eq. 1, all the projected data points in standard SC are the linear combination of unit vectors and the linear coefficients are in the interval of $[0,1]$. Hence, the bounding polygon is a convex polygon. □



(a) RadViz for 3D Case (b) RadViz for 4D Case (c) RadViz for 5D Case (d) RadViz for 6D Case

(e) Standard SC for 3D Case (f) Standard SC for 4D Case (g) Standard SC for 5D Case (h) Standard SC for 6D Case

Figure 3: The circumscribed circle and bounding polygon of RadViz and standard SC when plotting 3D, 4D, 5D, and 6D datasets.

Algorithm 1 Bounding polygon of standard SC

1: Rotate all the axes vectors so that the angle θ_1 of $\overrightarrow{v_1}$ is equal to 0;
2: Build V_1;
3: Calculate the starting point A;
4: Initialize the bounding polygon BP = [];
5: For $i = 1 : n$
6: Build V_i for $\overrightarrow{v_i}$;
7: If $i == n$
8: $V_{i+1} = V_1$;
9: Else
10: Build V_{i+1} for $\overrightarrow{v_i}$;
11: Find the corresponding $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$ for $\overrightarrow{v_i}$ as $\overrightarrow{v_a} \in (\overline{V_i} \cap V_{i+1})$ and $\overrightarrow{v_b} \in (V_i \cap \overline{V_{i+1}})$;
12: Calculate the order of $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$ according to Proposition 3.5;
13: Adding $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$ into BP in the calculated order;
14: End;

The convex polygon in RadViz is a regular polygon and the number of sides is equal to the number of dimensionality anchors [6]. However, in standard SC, though the convex polygon is also a regular polygon, the number of sides in this convex hull is not always equal to the number of dimensionality anchors. It depends on the dimensionality of the dataset.

Proposition 3.7. *Let R be the circumradius of the regular polygon, s be the side length, and N be the number of sides. We set the magnitude of axes vector to* 1. *If the dimensionality n is even, then $s = 2$ and $N = n$ and $R = 1/\sin(\pi/n)$. If n is odd, then $s = 1$ and $N = 2n$ and $R = 1/(2\sin(\pi/(2n)))$.*

Proof. By following the procedure described in Sect. 3.2.2, the obtained bounding polygon is formed by $2n$ vectors. However, in few cases, $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$ may be parallel. In Fig. 3, we illustrate the cases when using standard SC to plot 3D, 4D, 5D, and 6D datasets. In the 3D (Fig. 3(e)) and 5D (Fig. 3(g)) cases, $\overrightarrow{v_b}$ is not parallel with $\overrightarrow{v_a}$ or $\overrightarrow{v_{a-1}}$. Hence, when $n = 4\mathbb{Z}^+ - 1$ and $n = 4\mathbb{Z}^+ + 1$, the convex hull is a regular polygon with $N = 2n$. In the 4D case (Fig. 3(f)), though the obtained $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$ are not parallel, $\overrightarrow{v_b}$ is parallel with $\overrightarrow{v_{a-1}}$. While for the 6D case (Fig. 3(h)), the obtained $\overrightarrow{v_a}$ and $\overrightarrow{v_b}$ are parallel. Hence, when $n = 4\mathbb{Z}^+$ and $n = 4\mathbb{Z}^+ + 2$, the convex hull is a regular polygon with $N = n$ and the side length $s = 2$. □

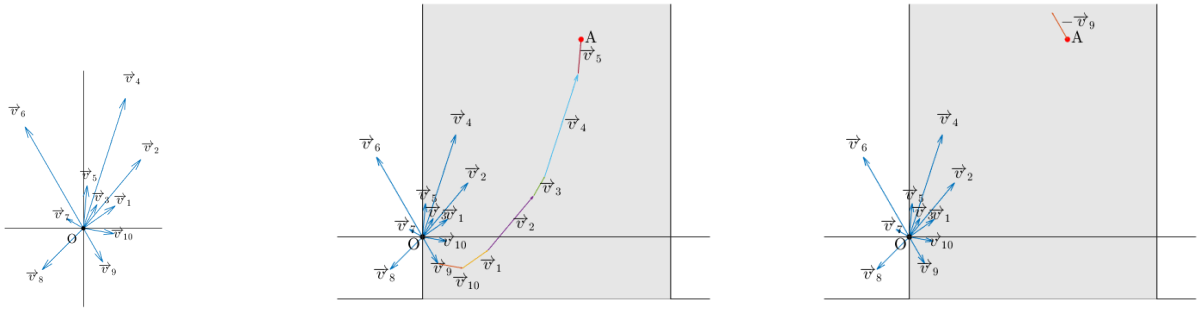
One issue to note is that the vertices of the bounding polygon obtained when $n = 4\mathbb{Z}^+$ and $n = 4\mathbb{Z}^+ + 2$ have different relative positions. In $n = 4\mathbb{Z}^+$ case, the vertices are not located along the axes vectors. While in $n = 4\mathbb{Z}^+ + 2$ case, vertices are located along the direction of the axes vectors.

4 EXPLORATION FOR BOUNDARY OF ARBITRARY SC

The boundaries of SC with arbitrary distribution configurations (AABB, bounding polygon, and minimum bounding box) are studied in this section. The magnitude L_i of vector $\overrightarrow{v_i}$ can be any non-negative value and $\theta_i \in [0, 2\pi)$. The study of the boundaries of arbitrary SC is more meaningful because finding the optimal configuration of axes vectors is the key point when using SC to perform data analysis.

4.1 Axis-aligned Bounding Box

For arbitrary SC, the magnitude of axes vectors $\overrightarrow{v_i}$ may not be of unit length. Hence, L_i should be considered when calculating the AABB.



(a) SC cases (b) Determine the starting point A (c) Determine the order of axes vectors

Figure 4: Determining the bounding polygon of SC. In (a), the configuration of SC is illustrated. Then the starting point A is confirmed as shown in (b). The third step is to confirm the order of axes vectors. In (c), from the starting point A, the first axes vector is confirmed and plotted.

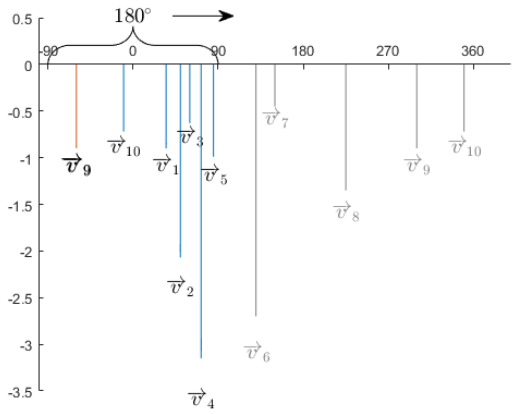


Figure 5: Determining the bounding polygon of SC by visualizing the magnitude and angle of each axes vector with a biplot map.

Proposition 4.1. *Let* $[X_{\max}, X_{\min}, Y_{\max}, Y_{\min}]$ *be the right, left, top, and bottom side of the AABB which contain the whole SC plotting. In arbitrary SC when plotting n-dimensional dataset, the AABB coordinates are*

$$\begin{cases} X_{\max} = \sum_{i\in I} L_i \cdot \cos\theta_i & with \quad I = \{i \in \mathbb{N} | \cos\theta_i \geq 0\} \\ X_{\min} = \sum_{i\in I} L_i \cdot \cos\theta_i & with \quad I = \{i \in \mathbb{N} | \cos\theta_i \leq 0\} \\ Y_{\max} = \sum_{i\in I} L_i \cdot \sin\theta_i & with \quad I = \{i \in \mathbb{N} | \sin\theta_i \geq 0\} \\ Y_{\min} = \sum_{i\in I} L_i \cdot \sin\theta_i & with \quad I = \{i \in \mathbb{N} | \sin\theta_i \leq 0\} \end{cases}, \quad (7)$$

where $\mathbb{N} = \{1, \cdots, n\}$.

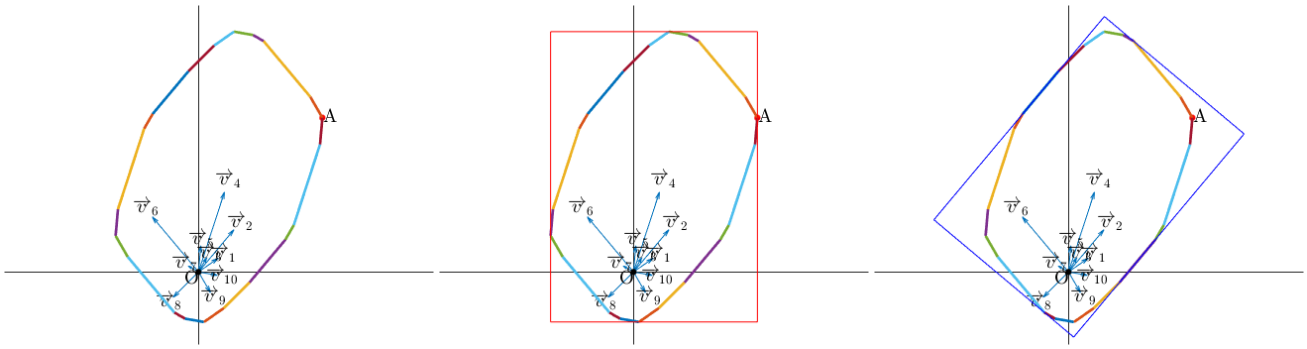
Proof. By following a similar strategy as in Proposition 3.1, all vectors $\overrightarrow{v_i}$ are projected onto the x-axis when calculating $X_{\max}$. When $\cos\theta_i > 0$, set the corresponding value $d'_{i,j}$ to one. Otherwise, set the $d'_{i,j}$ to zero. As a matter of fact, Eq. 7 is the generalized expression of Eq. 5. □

In arbitrary SC, the configuration may not be symmetric. Hence, the high-dimensional data point $d'_j = \{1, \ldots, 1\}$ may not be projected to the center of SC. In this case, Corollary 3.2 is invalid for arbitrary SC. Hence, the AABB of standard SC can be treated as a special case of the arbitrary SC's AABB.

4.2 Bounding Polygon

The algorithm to determine the bounding polygon of arbitrary SC is described here. Compared with standard SC cases, the bounding polygon determination of the arbitrary SC is much more complex. For arbitrary SC case, there are two main parts: determine the position of the starting point A and determine the order of axes vectors.

- Determine the position of the starting point A. The starting point is the place where we start to plot the bounding polygon. A configuration of an arbitrary SC plotting 10D dataset is shown in Fig. 4(a) and the corresponding axes vectors are also plotted in a biplot as shown in Fig. 5. We use 0 *rad* as the initial orientation. The starting point A is the farthest data point along the direction of 0 *rad*. The position of starting point A is equal to the sum of all the axes vectors in the shadow area in Fig. 4 (b). These axes vectors $V_{initial}$ have a positive value when projecting into the direction of 0 *rad*. This procedure can also be visualized in the biplot in Fig. 5. With the initial orientation 0 *rad*, axes vectors that are located in an interval of $[-\pi/2, \pi/2]$ are the related vectors. In this case, to build the starting point A, $V_{initial} = \{\overrightarrow{v_9}, \overrightarrow{v_{10}}, \overrightarrow{v_1}, \overrightarrow{v_2}, \overrightarrow{v_3}, \overrightarrow{v_4}, \overrightarrow{v_5}\}$ are used as shown in Fig. 4 (b).
- Determine the order of axes vectors one by one. As the bounding polygon is formed by axes vectors, the order is very important. Starting from point A, $\overrightarrow{v_6}$ and $\overrightarrow{v_9}$ are selected to compare their influence on the initial orientation $0rad$. Similar to Proposition 3.5, we compare the value of $(\theta_6 - \pi/2)$ and $(\theta_9 - 3\pi/2)$. In this case, $\theta_9 - 3\pi/2 < \theta_6 - \pi/2$, then the bounding polygon starts from $(-\overrightarrow{v_9}, \cdots)$ as shown in Fig. 4



(a) With bounding polygon (b) With axis-aligned bounding box (c) With minimum bounding box

Figure 6: The bounding polygon, the axis-aligned bounding box, and the minimum bounding box of the 10D SC plots.

(c). Another explanation is based on the biplot in Fig. 5. At the very beginning, the interval of $[-\pi/2, \pi/2]$ is centered at the initial orientation 0 *rad* to get the starting point A. Then to determine the order of the axes vectors, we can imagine the movement of the interval along the positive direction of the x-axis. In the initial moment, only $\overrightarrow{v_9}$, $\overrightarrow{v_{10}}$, $\overrightarrow{v_1}$, $\overrightarrow{v_2}$, $\overrightarrow{v_3}$, $\overrightarrow{v_4}$, and $\overrightarrow{v_5}$ are enclosed in the interval. During the movement,

- if an axes vector $\overrightarrow{v_a}$ breaks from the interval, then the bounding polygon will have this axes vector but with the opposite direction $-\overrightarrow{v_a}$. In this case, the bounding polygon will have $(-\overrightarrow{v_9}, \cdots)$.
- if an axes vector $\overrightarrow{v_a}$ falls into the interval, then the bounding polygon will have this axes vector with the same direction $\overrightarrow{v_a}$. In this case, the bounding polygon will have $(-\overrightarrow{v_9}, \overrightarrow{v_6}, \cdots)$.
- if an axes vector $\overrightarrow{v_a}$ breaks from the interval and another axes vector $\overrightarrow{v_b}$ falls into the interval at the same time, then the bounding polygon will have $-\overrightarrow{v_a}$ and $\overrightarrow{v_b}$ together. In this case, the order of $-\overrightarrow{v_a}$ and $\overrightarrow{v_b}$ has no influence on the bounding polygon as they are collinear.

The interval keeps moving until the bounding polygon has formed a closed path and the ending point has the same position with starting point A.

Proposition 4.2. *When transforming high-dimensional data point $d_j^{'}$ to $p_j^{SC}(x_j, y_j)$, the bounding polygon of arbitrary SC is a convex polygon.*

Proof. The proof of this proposition is the same as that in Proposition 4.2. Though the configuration of the SC has changed, a projected data point is still the linear combination of these axes vectors. Hence, the bounding polygon for any SC configuration is still convex. □

4.3 Minimum Bounding Box

The straight forward application of the bounding polygon is to find the minimum bounding box [29] enclosing the whole range of SC and then to rotate the whole view so as to maximize the display view of an arbitrary SC plotting. We follow the algorithm proposed in Sect. 4.1 and Sect. 4.2 to calculate the boundary of the 10D case as illustrated in Fig. 4 (a). The obtained bounding polygon and the axis-aligned bounding box are displayed in Fig. 6 (a) and (b). With the axis-aligned bounding box, a user can have the global view of the whole plotting. However, in some cases, the configuration of SC may lead to an odd bounding polygon. In such cases, the ratio of the bounding polygon area and the AABB area is very low. Hence, based on the obtained bounding polygon, we can calculate the relevant minimum bounding box to maximize the space utilization.

5 BOUNDARY-DRIVEN APPLICATION

5.1 SC with Boundaries

For standard SC, the relationship between magnitude and circumradius is described in Proposition 3.7. Based on this relationship, we fix the circumscribed circle radius of standard SC as the unit, then the magnitude of each axes vector can be obtained.

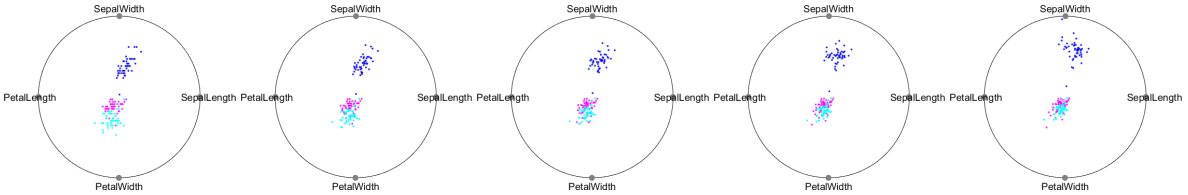
Let R be the circumradius of the bounding polygon of standard SC and we set $R = 1$. If the dimensionality n is even, then $L_i = \sin(\pi/n)$. If n is odd, then $L_i = 2\sin(\pi/(2n))$.

Hence, the expression of the standard SC in Eq. 1 can be modified as:

$$p_j^{SC-Even}(x_j, y_j) = \left(\sin(\pi/n) \cdot \sum_{i=1}^{n}(d_{i,j}^{'} \cdot \cos\theta_i),\ \sin(\pi/n) \cdot \sum_{i=1}^{n}(d_{i,j}^{'} \cdot \sin\theta_i)\right),$$

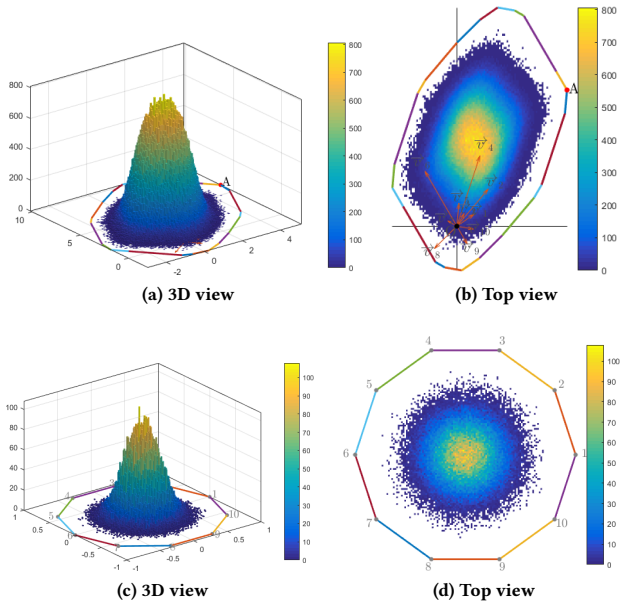
$$p_j^{SC-Odd}(x_j, y_j) = \left(2\sin(\pi/2n) \cdot \sum_{i=1}^{n}(d_{i,j}^{'} \cdot \cos\theta_i),\ 2\sin(\pi/2n) \cdot \sum_{i=1}^{n}(d_{i,j}^{'} \cdot \sin\theta_i)\right).$$

For arbitrary SC, the width and the height of the bounding box may not be the same. Hence, we set the larger one to one. In this case, the expression of the arbitrary SC in Eq. 1 can be modified as:



(a) $c = 0$ (b) $c = 0.25$ (c) $c = 0.50$ (d) $c = 0.75$ (e) $c = 1$

Figure 7: The smooth transition between standard SC and RadViz with fixed circumradius on the 'IRIS' dataset.



(a) 3D view (b) Top view

(c) 3D view (d) Top view

Figure 8: We plot 100,000 uniformly sampled 10-dimensional data points in SC with general configuration (a, b) and standard configuration (c, d) respectively to demonstrate the clumping effect.

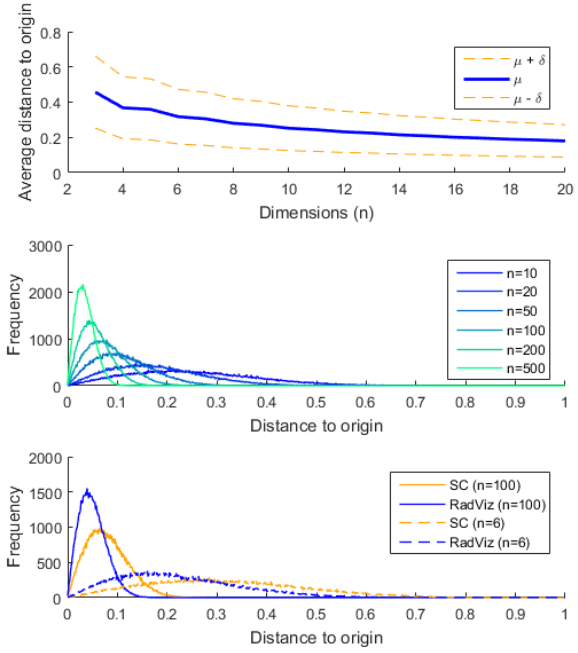


Figure 9: (Top) The average distance of all the projected data points to the origin decreases as the dimensionality increases, which means that the clumping effect becomes more obvious with high dimensional data. (Middle) The distribution of the distance using 10D, 20D, 50D, 100D, 200D, and 500D datasets [6]. (Bottom) Compared with the distribution of data point distance of RadViz, SC's clumping effect is less obvious, but both of them are increasing in dimensionality [23].

$$p_j^{SC}(x_j, y_j) = \left(\frac{\sum_{i=1}^{n}(d'_{i,j} \cdot L_i \cdot \cos\theta_i)}{max(X_{max} - X_{min}, Y_{max} - Y_{min})}, \frac{\sum_{i=1}^{n}(d'_{i,j} \cdot L_i \cdot \sin\theta_i)}{max(X_{max} - X_{min}, Y_{max} - Y_{min})}\right).$$

5.2 Association with RadViz

RadViz is often similarly compared with SC in recent works [18, 23]. The key difference between them is the non-linear factor as shown in Eq. 3. However, the sizes of these two kinds of plotting are not considered. In their comparison, the magnitude of axes vectors in both RadViz and SC are fixed to be one, i.e. $L_i = 1$. In standard SC, as shown in Sect. 3.1 and Sect. 3.2, if the magnitude of axes vectors is fixed, then the range of standard SC has a positive correlation with the dataset dimensionality n. Meanwhile, in RadViz, all the projected data points are located within the convex hull formed by the dimensionality anchors while all dimensionality anchors are located on a circle with a unit radius. As they did not consider the range of SC, they just use a view that can cover all the projected data points. Hence, it is not fair to compare a plotting which has a whole range with another plotting which only has partial range.

We fix the circumscribed circle radius of both RadViz and standard SC in order to compare these two plottings fairly. For RadViz,

the radius of the circumcircle for all the dimensionality is the same. So no change will be made in the expression of RadViz in Eq. 2. Based on the same circumradius, the configurations of RadViz and standard SC when plotting 3D, 4D, 5D, and 6D datasets are compared in Fig. 3. The bounding polygons of RadViz and standard SC have a larger difference in lower dimensionality cases. With increasing dimensionality, the bounding polygons of both RadViz and standard SC approach to the circumscribed circle.

The general projective maps (GPM) designed by Lehmann and Theisel is a linear interpolation between RadViz and SC [18]. However, the expression in [18] does not consider the range difference between RadViz and SC. Here, we reorganize the definition of GPM by combining Eq. 1 and Eq. 2 yielding the expression is

$$p_j^{GPM}(x_j, y_j) = \left(\frac{\sum_{i=1}^{n}(d'_{i,j} \cdot L_i \cdot \cos\theta_i)}{(1-c)\cdot 1 + c \cdot \sum_{i=1}^{n} d'_{i,j}}, \frac{\sum_{i=1}^{n}(d'_{i,j} \cdot L_i \cdot \sin\theta_i)}{(1-c)\cdot 1 + c \cdot \sum_{i=1}^{n} d'_{i,j}}\right), \tag{8}$$

where $c \in [0, 1]$. If $c = 1$, then GPM equals RadViz. Otherwise if $c = 0$, then it gives the SC. According to Eq. 8, the circumradius of GPM is set to be L_i and fixed along with different c. In this case, the smooth transition between RadViz and standard SC can be correctly expressed. As an example, Fig. 7 illustrates the smooth transition on the 'IRIS' dataset.

5.3 Clumping Effect Analysis

The determination of SC boundaries contributes to the feature analysis in which a global view is required. One case is the clumping effect of SC plotting. The clumping effect is known as the effect that forces data points towards the center of the plotting. The clumping problem of radial visualization method was first observed and analyzed and the effect of diametrically opposed dimensional anchors in RadViz is believed to be the main reason [6]. However, they did not analyze the clumping effect under different dimensionality and did not compare the clumping effect between RadViz and standard SC plots. The comparative study conducted by Rubio-Sánchez et al. [23] illustrated the comparison when plotting 6D and 100D datasets, but it did not consider the range change of SC along with the dimensionality changes. Here, we analyze the clumping effect of the standard SC under different dimensionality and fairly compare the clumping effect on RadViz and standard SC. Fig. 8 illustrates the clumping effect of SC with an arbitrary configuration and a standard configuration, respectively. The clumping effect can be clearly observed in these two kinds of SC configuration and the configuration difference has little effect on the clumping effect.

A quantitative analysis on the clumping effect on standard SC is conducted. The average distances of projected data points to the origin and the distributions of distances to the origin under difference dimensions are calculated in order to provide different views on the clumping effect of standard SC as shown in Fig. 9. The circumscribed circle radius is set to be one for different dimensions. The observation from Fig. 9 (Middle) is that with increasing dimensionality, the peak of the plotting is skewing to the left which indicates that an increasing number of data points are clumping towards the origin in standard SC. The residual space is sparsely populated, but it does not mean that it is empty.

The difference between our analysis and the previous result on the clumping effect of SC [23] is twofold. Firstly, the average distance decreases as the number of dimensionality increases in SC in our analysis while an inverse result was obtained in [23]. The previous analysis did not consider the range increase of SC as the dimensionality increases. Hence, the incorrect conclusion was obtained. Secondly, the comparison between the RadViz's clumping effect and the SC's clumping effect is different. In our analysis, the clumping effect of both RadViz and SC are more comprehensive for high-dimensional dataset as shown in Fig. 9 (Bottom). However, with the missing of the global view, the previous analysis failed to fairly compare the performance of RadViz and SC.

6 DISCUSSION AND CONCLUSION

Conclusion. This paper proposed boundary determination algorithms for both standard and arbitrary SC configurations. These boundaries provide users and researchers a global view on the SC plots that has never been done before. Now, it is possible for SC to plot subsequent data points without an additional operation to adjust the view and features such as clumping effect can also be analyzed. Moreover, equipped with the boundaries, the clumping effect of standard SC and arbitrary SC can be clearly visualized which can provide a clearer understanding on the plot. It is believed that the better understanding on the plots and the features will be helpful for further concrete numerical analysis.

Verification of the SC Boundary. The verification of the SC plots boundary, especially bounding polygon, can be conducted by many methods. One is to find the object-oriented minimum bounding boxes for each orientation and then the union of all these object-oriented minimum bounding box is the bounding polygon. This is due to the convexity of the boundary polygon as proved previously.

Computational Complexity. The computational complexity of our proposed algorithms is low. For the bounding polygon determination of arbitrary SC configuration, the result can be obtained in $O(n)$ time. Meanwhile, the minimum bounding box can be calculated in linear time. Hence, it is possible for the SC plots to dynamically adjust the view orientation to maximize the space utilization.

Future Work. The bounding polygon of SC can not only provide a better view, but also provide additional information. It is possible to give special meaning to the boundary edge or utilize the distance between the origin and the boundary. This is an interesting research direction to visualize continuous datasets, such as time-series data and to find the status changes.

ACKNOWLEDGMENTS

This work was conducted within the Rolls-Royce@NTU Corporate Lab with support from the National Research Foundation (NRF) Singapore under the Corp Lab@University Scheme. We also acknowledge the support by MOE grant RGC 2017-T1-001-053, Singapore.

REFERENCES

[1] Georgia Albuquerque, Martin Eisemann, Dirk J Lehmann, Holger Theisel, and Marcus Magnor. 2010. Improving the visual analysis of high-dimensional datasets

using quality measures. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*. IEEE, 19–26.

[2] Almir Olivette Artero and Maria Cristina Ferreira de Oliveira. 2004. Viz3d: Effective exploratory visualization of large multidimensional data sets. In *Computer Graphics and Image Processing, 2004. Proceedings. 17th Brazilian Symposium on*. IEEE, 340–347.

[3] Almir Olivette Artero, Maria Cristina Ferreira de Oliveira, and Haim Levkowitz. 2006. Enhanced high dimensional data visualization through dimension reduction and attribute arrangement. In *Information Visualization, 2006. IV 2006. Tenth International Conference on*. IEEE, 707–712.

[4] Enrico Bertini, Luigi Dell'Aquila, and Giuseppe Santucci. 2005. Springview: Cooperation of radviz and parallel coordinates for view optimization and clutter reduction. In *Coordinated and Multiple Views in Exploratory Visualization, 2005.(CMV 2005). Proceedings. Third International Conference on*. IEEE, 22–29.

[5] Keke Chen and Ling Liu. 2006. iVIBRATE: Interactive visualization-based framework for clustering large datasets. *ACM Transactions on Information Systems (TOIS)* 24, 2 (2006), 245–294.

[6] Karen Daniels, Georges Grinstein, Adam Russell, and Mason Glidden. 2012. Properties of normalized radial visualizations. *Information Visualization* 11, 4 (2012), 273–300.

[7] Mark De Berg, Marc Van Kreveld, Mark Overmars, and Otfried Cheong Schwarzkopf. 2000. Computational geometry. In *Computational geometry*. Springer, 1–17.

[8] Ronald A Fisher. 1936. The use of multiple measurements in taxonomic problems. *Annals of eugenics* 7, 2 (1936), 179–188.

[9] Jacob Goldberger, Geoffrey E Hinton, Sam T Roweis, and Ruslan R Salakhutdinov. 2005. Neighbourhood components analysis. In *Advances in neural information processing systems*. 513–520.

[10] Patrick Hoffman, Georges Grinstein, Kenneth Marx, Ivo Grosse, and Eugene Stanley. 1997. DNA visual and analytic data mining. In *Visualization'97., Proceedings*. IEEE, 437–441.

[11] Patrick Hoffman, Georges Grinstein, and David Pinkney. 1999. Dimensional anchors: a graphic primitive for multidimensional multivariate information visualizations. In *Proceedings of the 1999 workshop on new paradigms in information visualization and manipulation in conjunction with the eighth ACM internation conference on Information and knowledge management*. ACM, 9–16.

[12] Amin Ibrahim, Shahryar Rahnamayan, Miguel Vargas Martin, and Kalyanmoy Deb. 2016. 3D-RadVis: visualization of pareto front in many-objective optimization. In *Evolutionary Computation (CEC), 2016 IEEE Congress on*. IEEE, 736–745.

[13] Alfred Inselberg. 2009. *Parallel coordinates*. Springer.

[14] Eser Kandogan. 2000. Star coordinates: A multi-dimensional visualization technique with uniform treatment of dimensions. In *Proceedings of the IEEE Information Visualization Symposium*, Vol. 650. 22.

[15] Eser Kandogan. 2001. Visualizing multi-dimensional clusters, trends, and outliers using star coordinates. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 107–116.

[16] Yehuda Koren and Liran Carmel. 2003. Visualization of Labeled Data Using Linear Transformations.. In *INFOVIS*. Citeseer.

[17] Dirk J Lehmann and Holger Theisel. 2013. Orthographic star coordinates. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2615–2624.

[18] Dirk J Lehmann and Holger Theisel. 2016. General projective maps for multidimensional data projection. In *Computer Graphics Forum*, Vol. 35. Wiley Online Library, 443–453.

[19] Lenka Nováková and Olga Štepanková. 2009. Radviz and identification of clusters in multidimensional data. In *Information Visualisation, 2009 13th International Conference*. IEEE, 104–109.

[20] Lenka Nováková and Olga Štěpánková. 2011. Visualization of trends using RadViz. *Journal of Intelligent Information Systems* 37, 3 (2011), 355.

[21] Karl Pearson. 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2, 11 (1901), 559–572.

[22] Andy Pryke, Sanaz Mostaghim, and Alireza Nazemi. 2007. Heatmap visualization of population based multi objective algorithms. In *International Conference on Evolutionary Multi-Criterion Optimization*. Springer, 361–375.

[23] Manuel Rubio-Sánchez, Laura Raya, Francisco Diaz, and Alberto Sanchez. 2016. A comparative study between RadViz and Star Coordinates. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 619–628.

[24] Manuel Rubio-Sánchez and Alberto Sanchez. 2014. Axis calibration for improving data attribute estimation in star coordinates plots. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 2013–2022.

[25] Jahangheer S Shaik and Mohammed Yeasin. 2006. Visualization of high dimensional data using an automated 3D star co-ordinate system. In *Neural Networks, 2006. IJCNN'06. International Joint Conference on*. IEEE, 1339–1346.

[26] John Sharko, Georges Grinstein, and Kenneth A Marx. 2008. Vectorized radviz and its application to multiple cluster datasets. *IEEE transactions on Visualization and Computer Graphics* 14, 6 (2008).

[27] Marco Soldati, Mario Doulis, and Andre Csillaghy. 2007. SphereViz-Data Exploration in a Virtual Reality Environment. In *Information Visualization, 2007. IV'07. 11th International Conference*. IEEE, 680–683.

[28] Holger Theisel and Matthias Kreuseler. 1998. An enhanced spring model for information visualization. In *Computer Graphics Forum*, Vol. 17. Wiley Online Library, 335–344.

[29] Godfried T Toussaint. 1983. Solving geometric problems with the rotating calipers. In *Proc. IEEE Melecon*, Vol. 83. A10.

[30] Tran Van Long and Lars Linsen. 2011. Visualizing high density clusters in multidimensional data using optimized star coordinates. *Computational Statistics* 26, 4 (2011), 655.

[31] Yan-Chao Wang, Qian Zhang, Feng Lin, Chi-Keong Goh, Xuan Wang, and Hock-Soon Seah. 2017. Histogram equalization and specification for high-dimensional data visualization using RadViz. In *Proceedings of the Computer Graphics International Conference*. ACM, 15.

[32] Kilian Q Weinberger and Lawrence K Saul. 2009. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research* 10, Feb (2009), 207–244.

[33] Germain Garcia Zanabria, Luis Gustavo Nonato, and Erick Gomez-Nieto. 2016. iStar (i*): An interactive star coordinates approach for high-dimensional data exploration. *Computers & Graphics* 60 (2016), 107–118.