A comparative study between RadViz and Star Coordinates

Manuel Rubio-Sánchez, Laura Raya, Francisco Díaz and Alberto Sanchez

Abstract—RadViz and star coordinates are two of the most popular projection-based multivariate visualization techniques that arrange variables in radial layouts. Formally, the main difference between them consists of a nonlinear normalization step inherent in RadViz. In this paper we show that, although RadViz can be useful when analyzing sparse data, in general this design choice limits its applicability and introduces several drawbacks for exploratory data analysis. In particular, we observe that the normalization step introduces nonlinear distortions, can encumber outlier detection, prevents associating the plots with useful linear mappings, and impedes estimating original data attributes accurately. In addition, users have greater flexibility when choosing different layouts and views of the data in star coordinates. Therefore, we suggest that analysts and researchers should carefully consider whether RadViz's normalization step is beneficial regarding the data sets' characteristics and analysis tasks.

Index Terms—RadViz, Star coordinates, Exploratory data analysis, Cluster analysis, Classification, Outlier detection.

Many data visualization methods can be understood as dimensionality reduction techniques, which define procedures that map numerical multivariate data onto a low-dimensional observable display in order to represent them graphically. In this paper we evaluate and compare two well-known and closely related dimensionality reduction projection techniques that arrange and depict numerical variables in radial layouts, while representing high-dimensional data samples as points on low-dimensional plots. The first method is star coordinates (SC) [19, 20], which produces linear mappings by computing linear combinations of a set of low-dimensional vectors that represent radial axes. The second technique is RadViz [14, 13, 7], which instead generates nonlinear mappings of high-dimensional data onto a plane by modeling a physical spring system where the variables constitute anchor points. These techniques have been applied in fields as diverse as bioinformatics, engineering, finance, or nutrition; and have proven to be useful for several exploratory data analysis tasks, including cluster structure discovery, outlier and trend detection, feature extraction, or decision support tasks.

Despite the different design motivations for the methods, formally, the main difference between them consists of a normalization step inherent in RadViz, which introduces nonlinearities. The contribution of this paper is an objective comparison of both methods when performing diverse exploratory analysis tasks, making special emphasis on the effect of RadViz's extra normalization step. In particular, we show that it can be useful when the data is sparse, i.e., when data samples contain only a few (related) attributes that are considerably larger than the rest. However, in general it can hamper several exploratory analysis tasks since it: (1) introduces distortions in the plots that affect the distances between the mapped points, (2) can encumber outlier detection, (3) prevents computing configurations of radial axes that lead to powerful and well-known linear plots, which can be obtained automatically, and can be useful for tasks such as classification or feature extraction, and (4) impedes estimating original data attributes accurately. Moreover basic SC visualizations may even provide better results than sophisticated approaches developed for RadViz. Finally, we argue that users have a greater flexibility when choosing different layouts and views of the data in SC.

The rest of the paper is organized as follows. Section 1 provides an overview of the related work. Section 2 describes SC and RadViz in detail, and shows the equivalence between them. Section 3 analyzes the effect on the plots caused by the nonlinear step in RadViz, while Sec. 4 analyzes the ability of the methods to detect two types of outliers. Section 5 shows the usefulness of obtaining automatic configurations of axis vectors in SC associated to well-known linear mappings, while Sec. 6 examines how users can recover attribute values in SC and RadViz. Finally, Sec. 7 compares the methods' plots when using arbitrary layouts, and Sec. 9 presents a discussion.

1 Related Work

RadViz is arguably the archetype of radial projection approaches. Numerous papers have proposed ideas in order to enhance or extend the method (see [15, 37, 36, 28, 33]). The research has often focused on selecting layouts for the anchor points automatically in order to perform or optimize tasks such as cluster or class separation, or outlier detection (see [26, 3, 21, 1, 5, 34]). These approaches propose complex heuristics in order to accomplish the analysis tasks, sometimes posing NP-hard problems (e.g., for sorting the anchor points around a circle). Finally, researchers have also proposed to combine RadViz with other approaches [4, 29], and have selected it to represent radial methods when comparing and analyzing different types of visualization approaches [12, 30, 42].

Regarding SC, research has also focused on automatically finding configurations of axis vectors in order to accomplish or optimize data analysis tasks (see [35, 41, 38]). Again, the proposed approaches typically rely on heuristics that can be complex or time consuming. Recently it has been shown that SC plots can be enhanced by configuring the axes in order to produce orthogonal projections [23]. In addition, when the data is centered it is possible to recover original data attributes more accurately, which altogether leads to more faithful representations of the data [32].

Dimensionality reduction methods can be categorized as linear or nonlinear. Linear techniques map high-dimensional points onto a low-dimensional space by a simple matrix-vector product, and have been studied extensively. Popular methods include principal component analysis (PCA) [17], statistical biplots [9], linear discriminant analysis (LDA) [27], or projection pursuit [18]. In addition, over the past decade, researchers in the field of machine learning have developed powerful metric learning methods for classification, but which also provide linear maps that can be applied in order to represent the data in 2 or 3 dimensions [10, 43]. We will show in Sec. 5 that it is possible to select axis vectors in SC in order to reproduce these plots.

In contrast, modern nonlinear techniques construct sophisticated mappings in an attempt to represent data that lies on a lower-dimensionality manifold as faithfully as possible (see [39, 31, 25]). These methods take into account relationships between the data samples (e.g., building a neighborhood graph) to generate useful low-dimensional representations. However, unlike these methods, RadViz defines a general nonlinear operation that does not consider relationships between the data. Thus, its mapping is essentially different, and

- *Manuel Rubio-Sánchez is with URJC. E-mail: manuel.rubio@urjc.es.*
- *Laura is with U-tad. E-mail: laura.raya@gmail.com.*
- *Francisco Díaz is with UPM. E-mail: fdiaz@eui.upm.es.*
- *Alberto Sanchez is with URJC and CCS. E-mail: alberto.sanchez@urjc.es.*

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication xx Aug. 2015; date of current version 25 Oct. 2015. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

is used for other purposes.

2 Description of the Methods

This section briefly summarizes SC and RadViz (introducing the main notation used throughout the paper), and describes the conditions under which they are equivalent.

2.1 Star coordinates

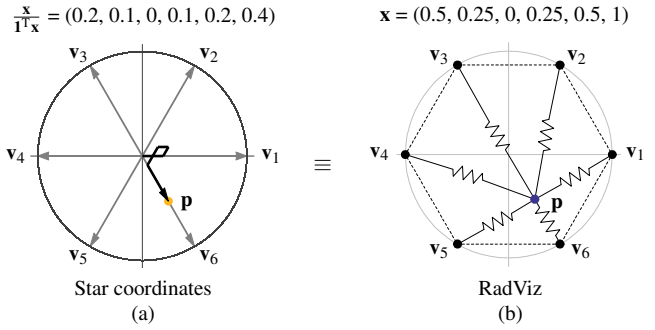


Fig. 1. Equivalence between SC and RadViz. Let $\{\mathbf{v}_i\}$ denote the axis vectors of a SC plot, shown in (a), as well as the set of anchor points of a RadViz model, illustrated in (b). The low-dimensional representation $\mathbf{p}$ of some data sample $\mathbf{x}$ with nonnegative attributes in RadViz is the same as that of $\mathbf{x}/(\mathbf{1}^T\mathbf{x})$ (i.e., $\mathbf{x}$ normalized so that the sum of its elements is 1) in SC, where $\mathbf{1}$ is the vector of all ones.

Star coordinates generates linear mappings from an n-dimensional data space onto a lower m-dimensional observable space ($m \leq 3$) in order to represent the data graphically. In particular, it constructs plots through a set of m-dimensional vectors $\mathbf{v}_i$, for $i = 1,\ldots,n$, with a common origin point that represent radial axes, where $\mathbf{v}_i$ is associated with the i-th data variable (see Fig. 1a). The low-dimensional embedding $\mathbf{p} \in \mathbb{R}^m$ of a data sample $\mathbf{x} \in \mathbb{R}^n$ is simply a linear combination of the vectors $\mathbf{v}_i$, where the linear coefficients correspond to the variable attributes of $\mathbf{x}$. Formally:

$$\mathbf{p} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n = \mathbf{V}^T\mathbf{x}, \tag{1}$$

where $\mathbf{V}$ is the $n \times m$ matrix whose rows are the vectors $\mathbf{v}_i$. The interpretation of the axis vectors is straightforward: the orientation determines the direction in which a variable increases, and the length specifies the amount of contribution of a particular variable in the resulting visualization, given that all variables have a similar scaling. Note that the mapping is linear since it consists of a matrix-vector product.

2.2 RadViz

RadViz implements a physical spring model metaphor. Similarly to SC, it uses a set of m-dimensional vectors $\mathbf{v}_i$, for $i = 1,\ldots,n$, which define anchor points of n springs. These, in turn, are connected to the low-dimensional representation $\mathbf{p}$ of some data sample $\mathbf{x} \in \mathbb{R}^n$ (see Fig. 1b). The attributes of $\mathbf{x}$, which must be nonnegative, determine the springs' stiffness, and the location of $\mathbf{p}$ is such for which the sum of the spring forces equals 0 (i.e., the point lies at an equilibrium position).

Formally, RadViz generates nonlinear projections according to:

$$\mathbf{p} = \frac{\sum_{i=1}^n x_i\mathbf{v}_i}{\sum_{i=1}^n x_i}, \tag{2}$$

where the data values x_i must be nonnegative (they are usually normalized so that the range of each variable is the $[0,1]$ interval). Note that the method's nonlinearity stems from the term in the denominator. Lastly, the zero vector is mapped onto the center of mass (i.e., mean or barycenter) of the anchor points.

The mapping entails the following properties: (1) the larger x_i is in comparison with the rest of values, the closer $\mathbf{p}$ will be to the anchor point $\mathbf{v}_i$, (2) samples whose data attributes are all the same get mapped onto the center of mass of the anchor points, (3) the point $\mathbf{p}$ is a convex combination of the anchor points $\mathbf{v}_i$ for which $x_i > 0$, and will therefore lie inside their convex hull, and (4) if the i-th element of a sample is the only nonzero one, then it will be mapped onto $\mathbf{v}_i$ (while samples with only two nonzero attributes get mapped to a point on the line segment between the associated anchor points). The converse statement of this last property is one of the most important features of RadViz regarding its usefulness. In order to ensure it, all of the anchor points must be different and form part of their convex hull. Therefore, in practice the anchor points are typically arranged around a circle.

2.3 Equivalence between RadViz and star coordinates

The only difference between both mappings resides in the denominator in (2) [28], except that RadViz requires the data to be nonnegative. In the following proposition we formally describe the conditions under which SC and RadViz are equivalent (see Fig. 1).

Proposition 1. *Let $\mathbf{V}$ be the $n \times m$ matrix whose rows consist of the anchor points of a RadViz model, as well as the axis vectors of a SC plot. The mapping $\mathbf{p} \in \mathbb{R}^m$ of an n-dimensional data point $\mathbf{x} \neq \mathbf{0}$, with nonnegative entries, is identical for both methods if the data is preprocessed in SC so that the sum of its elements is equal to 1.*

Proof. The preprocessing consists of dividing each attribute of a data sample by the sum of all of the attributes. Thus, due to (1) the mapping in SC is:

$$\mathbf{p}_{\text{SC}} = \mathbf{V}^T\left(\frac{\mathbf{x}}{\mathbf{1}^T\mathbf{x}}\right) = \frac{\sum_{i=1}^n x_i\mathbf{v}_i}{\sum_{i=1}^n x_i} = \mathbf{p}_{\text{RadViz}},$$

due to (2), where $\mathbf{1}^T$ is the (row) vector of all ones. □

Thus, the normalization step inherent in RadViz produces nonlinear mappings of the data onto the observable display, where we can consider that it applies SC's linear mapping, but to samples (with nonnegative entries) previously projected nonlinearly onto the unit $(n-1)$-simplex ($\mathbf{1}^T\mathbf{x} = 1$). Regarding the $\mathbf{0}$ vector, SC maps it onto the origin of the low-dimensional space, while RadViz projects it onto the center of mass of the anchor points (in practice it is often the origin as well, since the anchors are usually arranged uniformly around a circle).

3 Nonlinearity and Distortions

The geometrical properties of RadViz have been studied in depth recently in [7]. We now present a series of new results related to RadViz and SC, focusing on their difference.

3.1 Line segments and convex sets

In RadViz, line segments in the high-dimensional data space are mapped onto line segments on the plots [7], which, by definition, implies that convex sets also get mapped onto convex sets. This is also true for SC since it generates linear mappings, as we show in the following result.

Proposition 2. *In SC lines in the data space ($\mathbb{R}^n$) are mapped onto lines, or a single point (in the degenerate case), in the observable display ($\mathbb{R}^m$).*

Proof. Let $L = \{\mathbf{x} + \alpha\mathbf{u} | \alpha \in \mathbb{R}\}$ represent a line in the data space with some arbitrary direction $\mathbf{u} \in \mathbb{R}^n$, which passes through some arbitrary point $\mathbf{x} \in \mathbb{R}^n$. Additionally, let $\mathbf{V} \in \mathbb{R}^{n\times m}$ represent a matrix of axis vectors in SC. The linear mapping of L given the SC model is:

$$\{\mathbf{V}^T(\mathbf{x} + \alpha\mathbf{u}) | \alpha \in \mathbb{R}\} = \{\mathbf{V}^T\mathbf{x} + \alpha\mathbf{V}^T\mathbf{u} | \alpha \in \mathbb{R}\} = \{\mathbf{p} + \alpha\mathbf{w}) | \alpha \in \mathbb{R}\},$$

which is a line in the observable display with direction $\mathbf{w} \in \mathbb{R}^m$, that passes through $\mathbf{p} \in \mathbb{R}^m$. If $\mathbf{u} \in \mathcal{N}(\mathbf{V}^T)$, where $\mathcal{N}$ denotes nullspace, then $\mathbf{w} = \mathbf{0}$ and the line would get mapped onto the single point $\mathbf{p} = \mathbf{V}^T\mathbf{x}$. □

Moreover, since SC does not introduce nonlinearities, the low-dimensional embeddings of uniformly distributed points along a line segment in the data space will also be uniformly distributed on the SC plots, as we show in the following result.

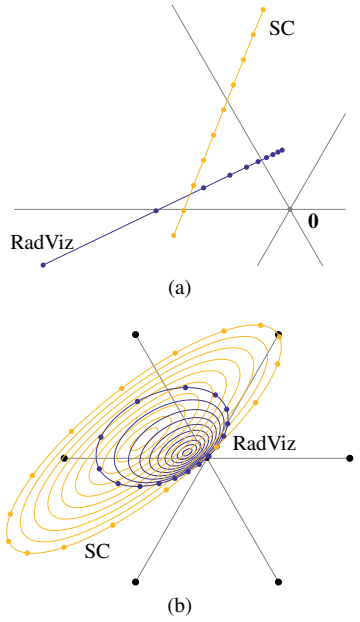


Fig. 2. Nonlinearities in RadViz. In (a) line segments are mapped onto line segments in both SC and RadViz. In SC points uniformly distributed along a line segment in the data space are mapped onto (light orange) points also uniformly distributed along a line segment on the display. However, this does not occur in RadViz, where (dark blue) points are concentrated towards an endpoint of the line segment. In (b) concentric circles are transformed into concentric ellipses when using SC. However, in RadViz the ellipses are no longer concentric. The plot also shows the mappings of points arranged uniformly on the largest concentric circle. The example uses 6 variables, where the anchor points (RadViz), or endpoints of the axis vectors (SC), represented by the darkest dots, are arranged in a regular configuration.

Proposition 3. *In SC points uniformly distributed along a line segment in the data space ($\mathbb{R}^n$) are mapped onto points also uniformly distributed along a line segment, or a single point (in the degenerate case), in the observable display ($\mathbb{R}^m$).*

Proof. Let $\mathbf{V} \in \mathbb{R}^{n\times m}$ represent a matrix of axis vectors in SC. Also, let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ represent the endpoints of a line segment in the data space, which we can define as $\alpha\mathbf{x} + (1-\alpha)\mathbf{y}$, for $0 \leq \alpha \leq 1$.

If $\mathbf{V}^{\mathrm{T}}\mathbf{x} = \mathbf{V}^{\mathrm{T}}\mathbf{y} = \mathbf{p}$ then all of the points in the line segment are mapped onto $\mathbf{p}$, since $\mathbf{V}^{\mathrm{T}}\big(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}\big) = \mathbf{V}^{\mathrm{T}}\mathbf{y} = \mathbf{p}$.

Otherwise, points uniformly distributed along the line segment from $\mathbf{x}$ to $\mathbf{y}$ will be mapped to points also uniformly distributed along the line segment from the low-dimensional representations of $\mathbf{x}$ and $\mathbf{y}$ in SC if and only if:

$$\mathbf{V}^{\mathrm{T}}\big(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}\big) = \theta\mathbf{V}^{\mathrm{T}}\mathbf{x} + (1-\theta)\mathbf{V}^{\mathrm{T}}\mathbf{y} \Rightarrow \alpha = \theta,$$

which is true in SC, since:

$$\begin{aligned}&\mathbf{V}^{\mathrm{T}}\big(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}\big)\\ &= \alpha\mathbf{V}^{\mathrm{T}}\mathbf{x} + (1-\alpha)\mathbf{V}^{\mathrm{T}}\mathbf{y} = \theta\mathbf{V}^{\mathrm{T}}\mathbf{x} + (1-\theta)\mathbf{V}^{\mathrm{T}}\mathbf{y} \Leftrightarrow \alpha = \theta.\end{aligned}$$

□

Thus, ignoring the degenerate cases, the shape of distributions of points along a line segment is preserved under the SC mapping. However, in RadViz uniformly distributed points along a line segment in the data space are no longer mapped onto points also uniformly distributed along a line segment in the observable display, as we show in the following result.

Proposition 4. *In RadViz points uniformly distributed along a line segment in the data space ($\mathbb{R}^n$) are not mapped onto points also uniformly distributed along a line segment in the observable display ($\mathbb{R}^m$).*

Proof. Let $\mathbf{V} \in \mathbb{R}^{n\times m}$ represent a matrix of anchor points in RadViz. Also, let $\mathbf{x}$ and $\mathbf{y}$ be points in $\mathbb{R}^n$. Points uniformly distributed along the line segment from $\mathbf{x}$ to $\mathbf{y}$ will be mapped to points also uniformly distributed along the line segment from the low-dimensional representations of $\mathbf{x}$ and $\mathbf{y}$ in RadViz if and only if:

$$\begin{aligned}\mathbf{V}^{\mathrm{T}}\left[\frac{\alpha\mathbf{x} + (1-\alpha)\mathbf{y}}{\mathbf{1}^{\mathrm{T}}(\alpha\mathbf{x} + (1-\alpha)\mathbf{y})}\right] &= \theta\mathbf{V}^{\mathrm{T}}\left(\frac{\mathbf{x}}{\mathbf{1}^{\mathrm{T}}\mathbf{x}}\right) + (1-\theta)\mathbf{V}^{\mathrm{T}}\left[\frac{\mathbf{y}}{\mathbf{1}^{\mathrm{T}}\mathbf{y}}\right]\\ &\Rightarrow \alpha = \theta.\end{aligned}$$

Multiplying both sides of the first equality by $(\mathbf{1}^{\mathrm{T}}\mathbf{x})(\mathbf{1}^{\mathrm{T}}\mathbf{y})$ we obtain:

$$\begin{aligned}&\left[\frac{\alpha(\mathbf{1}^{\mathrm{T}}\mathbf{x})}{\alpha(\mathbf{1}^{\mathrm{T}}\mathbf{x}) + (1-\alpha)(\mathbf{1}^{\mathrm{T}}\mathbf{y})}\right](\mathbf{V}^{\mathrm{T}}\mathbf{x})(\mathbf{1}^{\mathrm{T}}\mathbf{y})\\ +\ &\left[\frac{(1-\alpha)(\mathbf{1}^{\mathrm{T}}\mathbf{y})}{\alpha(\mathbf{1}^{\mathrm{T}}\mathbf{x}) + (1-\alpha)(\mathbf{1}^{\mathrm{T}}\mathbf{y})}\right](\mathbf{V}^{\mathrm{T}}\mathbf{y})(\mathbf{1}^{\mathrm{T}}\mathbf{x})\\ =\ &\theta(\mathbf{V}^{\mathrm{T}}\mathbf{x})(\mathbf{1}^{\mathrm{T}}\mathbf{y}) + (1-\theta)(\mathbf{V}^{\mathrm{T}}\mathbf{y})(\mathbf{1}^{\mathrm{T}}\mathbf{x}),\end{aligned}$$

which is satisfied if and only if:

$$\theta = \frac{\alpha\mathbf{1}^{\mathrm{T}}\mathbf{x}}{\alpha\mathbf{1}^{\mathrm{T}}\mathbf{x} + (1-\alpha)\mathbf{1}^{\mathrm{T}}\mathbf{y}}.$$

Thus, $\alpha \neq \theta$ in general. □

This creates nonlinear distortions since it concentrates the mapped points towards one of the endpoints (see Fig. 2a). The resulting plots can be misleading, since the relative distances between the data points are altered. Figure 2b shows the effect when mapping a set of concentric circles, which roughly illustrates how a two-dimensional (intrinsic dimensionality) cluster in the data space is mapped onto the display in RadViz. The plot shows a similar effect as in Fig. 2a, by mapping 16 data points arranged uniformly on the largest concentric circle.

3.2 Cluster representation

According to [6], the visual interpretation of the clusters is more difficult in RadViz than in SC. In order to back up this claim we present examples that show the effect of the nonlinearities introduced by RadViz when plotting clusters.

Figure 3 shows RadViz and SC mappings of a multivariate normal distribution in $\mathbb{R}^6$. In (a) the cluster of mapped points in the RadViz plot does not follow a normal distribution. This can be visually tested with the Q-Q plot in (b). In particular, note that for multivariate normal distributions the squared Mahalanobis distances from the points to their mean follow a χ^2_m distribution (with m degrees of freedom, where $m = 2$ is the dimensionality of the data). Thus, we have plotted the quantiles of these Mahalanobis distances with respect to the theoretical quantiles of the χ^2_2 distribution. It is clear that the distribution of Mahalanobis distances has heavier tails, which explains why several points on the RadViz plot lie far away from the cluster. In practice these points could be misleading since they could be mistaken for outliers, or for samples belonging to a different cluster. Alternatively, the SC plot in (c) shows a normally distributed cluster. In this case, it must be normally distributed, since an affine transformation of a multivariate normal is also normally distributed. Thus, the associated Q-Q plot in (d) shows points lined up along a straight 45° line.

RadViz's mapping also tends to clump embedded points close to the origin [7]. Fig. 4 compares the average distance from an embedded point to the origin in RadViz and SC, for 10^6 randomly chosen data points in $[0,1]^n$, and regular configurations of axis vectors and anchor points (with $\|\mathbf{v}_i\| = 1$). These average distances decrease as the number of variables increases in RadViz, while they increase in SC, as shown in (a). Nevertheless, since analysts can zoom in on the plots, the relative clumping of points can be analyzed by comparing the shapes of the distributions of distances to the origin. The graphic

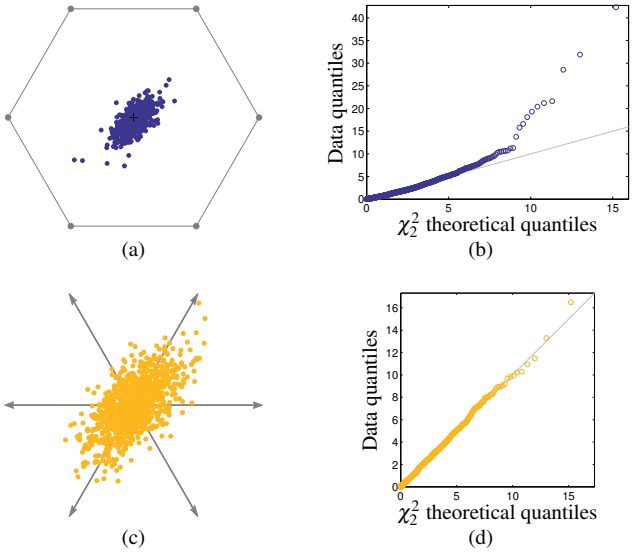


Fig. 3. Mappings of 1000 samples drawn from a multivariate normal distribution in $\mathbb{R}^6$. The plot in (a) corresponds to RadViz's mapping, where several points appear far away from the cluster, and could be misinterpreted as outliers. In (b) we have used a Q-Q plot to test whether the distribution of mapped points is normal. In this case, the graph clearly indicates that the mapped cluster does not follow a normal distribution. Instead, the SC plot in (c) does (and must, since the mapping is linear) show a normally distributed cluster. Thus, the points on the associated Q-Q plot in (d) appear along a straight 45° line.

in (b) shows line histograms of these distances for both methods using 6 and 100 variables. For a low number of variables the clumping effect is more pronounced in RadViz, since the distribution is more skewed to the left. However, the shapes of the distributions are very similar as n increases.

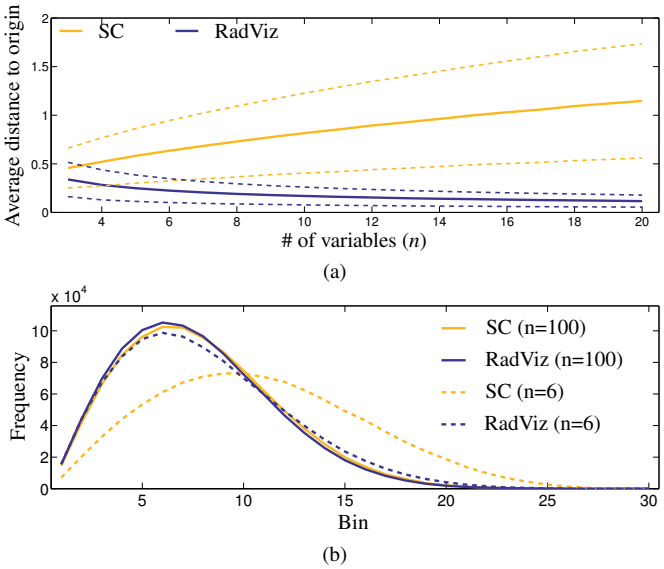


Fig. 4. Analysis of the clumping of points towards the origin in RadViz and SC, for regular configurations of variables. The average distance of a point to the origin in the display decreases as the number of variables increases in RadViz, while it increases in SC, as shown in (a). The plot also includes the mean ± one standard deviation (dashed lines). The graphic in (b) shows line histograms of the distances for both methods using 6 and 100 variables. The clumping effect is more pronounced in RadViz, but the shapes of the distributions are very similar for a large number of variables.

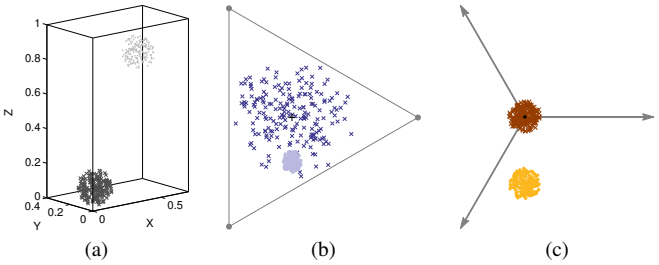


Fig. 5. Cluster size preservation in SC. A 3-dimensional toy data set described in [28] containing two ball-shaped clusters with equal radii and 200 points is shown in (a). The RadViz plot in (b) shows the clusters with different sizes, while in the SC plot in (c) the clusters only differ in their location, but not in their size.

In RadViz the appearance of a cluster in the low-dimensional plot not only depends on its shape and size, but also on where it is located in the data space. In particular, the closer it lies to the origin, the larger it will appear on the display. This does not occur in SC since it produces linear mappings, as we show in the following result.

Proposition 5. *Let C_1 be any particular set of data samples in the data space $\mathbb{R}^n$, and let $C_2 = \{\mathbf{x}+\mathbf{a} | \mathbf{x} \in C_1\}$, for some arbitrary $\mathbf{a} \in \mathbb{R}^n$ (i.e., C_1 after applying some translation $\mathbf{a}$). Additionally, let $\mathbf{V} \in \mathbb{R}^{n \times m}$ represent a matrix of axis vectors in SC. The low-dimensional embeddings of C_1 and C_2 on the SC plot will be identical up to a translation determined by the vector $\mathbf{V}^\mathrm{T}\mathbf{a}$.*

Proof. The set of low-dimensional embedded points of C_1 is $\{\mathbf{V}^\mathrm{T}\mathbf{x} | \mathbf{x} \in C_1\}$, while the corresponding set of mapped points for C_2 is $\{\mathbf{V}^\mathrm{T}(\mathbf{x}+\mathbf{a}) | \mathbf{x} \in C_1\} = \{\mathbf{V}^\mathrm{T}\mathbf{x} + \mathbf{V}^\mathrm{T}\mathbf{a} | \mathbf{x} \in C_1\}$. □

Figure 5 shows an example with a toy data set used in [28] (the supplemental material includes another example). It contains two ball-shaped clusters in $\mathbb{R}^3$ of radius 0.1, each with 200 points, as shown in (a). The darker one is centered near the origin at (0.1, 0.1, 0.1), while the lighter one is centered at (0.5, 0.2, 0.8). In RadViz the darker cluster appears larger than the lighter one, since it is located closer to the origin in the data space, as illustrated in (b). In order to avoid this issue, as well as the cluster overlap, [28] proposes to normalize the attribute values of the data in the interval $[a, 1]$, for some suitable $a > 0$. However, this heuristic is data dependent and does not solve the problem entirely. Additionally, since the normalization step in RadViz essentially eliminates the information regarding the ($\mathscr{L}_1$) distance from a data sample to the origin, [28] proposes to add a third dimension to the plots containing the Euclidean ($\mathscr{L}_2$) distance from the sample to the origin. Note, however, that simply using SC solves these issues regarding cluster size and overlap, as shown in (c).

3.3 Sparse data

RadViz has been used successfully in fields such as bioinformatics or biomedicine, where it is important to visualize the most expressed attributes (e.g., in microarray gene expression data) of data samples (see [13, 26, 33]). Its main advantage consists of the ability to map sparse data samples close to anchor points, and non-sparse data close to the origin. In this context we refer to sparse data as such for which one of the attributes values (or a few if their anchor points are close) is considerably greater than the rest, in which case RadViz would map the point close to the corresponding anchor. This type of data can also be created by procedures such as data flattening, where categorical variables are decomposed into sets of binary attributes (see [37]).

Figure 6 compares RadViz (a) and SC (b) regarding their ability to detect sparse data. The plots show the mappings of randomly generated data in $[0, 1]^{10}$, and sparse data (through lighter and larger colored dots). In the example, the sparse data consist of samples where one attribute value (in particular, the one associated with the darker anchor point or axis vector) is 20 times larger than maximum value of the remaining attributes of the sample. In SC the sparse data is plotted close

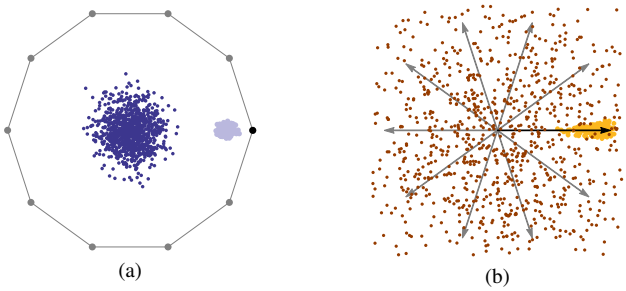


Fig. 6. Usefulness of RadViz for visualizing sparse data. The examples consist of a RadViz (a) and SC (b) plot of data in $[0,1]^{10}$ (1000 darker dots), and sparse data (200 lighter and larger dots). The sparse samples contain one attribute value (in particular, the one associated with the variable represented with a darker anchor point or axis vector) that is at least 20 times larger than the rest of attributes of the sample. Note that it is easier to distinguish these sparse data in RadViz.

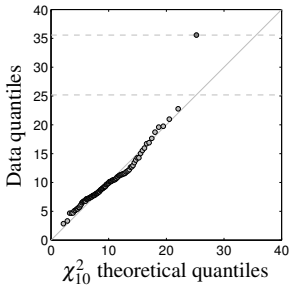


Fig. 7. Generation of a low likelihood data point, illustrated through a Q-Q plot. Firstly, N samples are drawn from a standard multivariate normal in an n-dimensional space, where the largest squared Mahalanobis distance to their mean is less or equal to the largest theoretical quantile ($F_n^{-1}(1-1/(2N)) = 25.19$, for $N = 100$ and $n = 10$), where F_n denotes the cumulative distribution function of a χ^2_n distribution (with n degrees of freedom). Finally, we generated an outlier, with entries in [-3,3], such that the squared Mahalanobis distance to the mean of the normal data was $F_n^{-1}(1-1/(100N))$, which is 35.56, for $N = 100$ and $n = 10$. Note that the outlier is clearly visible on the Q-Q plot.

to the tip of the corresponding axis vector, but non-sparse data may also be located in that region of the plot. Thus, it is easier to detect the sparse data in RadViz, since the mapped sparse points appear far away from non-sparse data.

Lastly, if the samples only contain one nonzero value, then SC would depict the data more faithfully than RadViz. In this case the mapped points would lie exactly on an axis, where they could be highlighted, and analysts could recover the nonzero value visually. In particular, it would be the distance from the mapped point to the origin, normalized by the length of the axis vector. In RadViz, the mapped point would be located at the corresponding anchor, and it would not be possible to recover the original nonzero value.

4 Outlier Detection

Evaluating the capability of a visualization method to detect outliers is nontrivial. On the one hand, there exist several definitions for outliers (see [28]). On the other hand, the binary decision of whether or not a sample is an outlier is often subjective, and may require domain knowledge [6]. In this paper we consider a common definition in which an outlier is a data sample that lies (probabilistically) far from the rest of the data. Note, however, that other authors have used different definitions when using RadViz or SC. For instance, an outlier can be understood as a data sample that is misclassified by classification methods [21], or that is not consistently grouped within the same set by different clustering algorithms [37].

In this paper we evaluate the probability of detecting two types of outliers using RadViz and SC, which are easily detected by classical statistical visualizations. In these experiments we have used sets of points sampled from a multivariate standard normal distribution, to which we have added an outlier. In order to ensure that the normally distributed data did not contain points that could be interpreted as outliers, we used the fact that the squared Mahalanobis distances from the points to their mean follow a χ^2_n distribution (see Sec. 3.2). In particular, we only considered data sets for which the largest squared Mahalanobis distance was less or equal to the largest theoretical quantile.

The first type of outlier consisted in modifying one (randomly chosen) attribute value of a sample (also selected at random) in the data set, by assigning to it an extreme value. We considered two different values depending on whether the variables were preprocessed in order to lie in the [0,1] interval. If so, we chose an extreme value of 1000. Thus, after the normalization the chosen attribute's value was 1 for the outlier, while for the rest of the data samples the value was very close to 0. However, when the data were not preprocessed the extreme value was 10. Note that the extreme values are exceedingly unlikely, and would therefore be easily detected in a box plot or histogram.

For the second type of outlier we included an additional sample that would have a low likelihood of belonging to the normally distributed data set, but would not contain extreme values. In particular, we chose a data sample, with attribute values inside the interval $[-3,3]$, such that the probability of observing another data sample with lower likelihood was $1/(100 \cdot N)$, where N is the cardinality of the normally distributed data set (roughly speaking, we would need to generate a data set 100 times larger in order to find a rarer sample). This sample is easily detected in Q-Q plots like those shown in Fig. 3. Figure 7 shows the process of generating the normal data and the low likelihood data point.

Lastly, given the SC or RadViz plots, we modeled two strategies users could take when searching for outliers. The first one consists of considering the farthest point from the center of mass of the mapped points. The second one involves finding the point that is farthest from any other mapped point. Note that both strategies can be simulated computationally, which allowed us to generate a large number of experiments. Figure 8 shows the average probabilities of detecting the introduced outlier using both strategies, over 10^5 trials, considering regular configurations of axis vectors and anchor points (the supplemental material includes results on random configurations, which are similar, but poorer). The normal data set contained $N = 100$ samples (results were essentially the same for other choices of N). We compared results for SC, Radviz, and the variant described in [28] for outlier detection, denoted as RadViz'. The graphs show that outliers generated through extreme values are detected more frequently using SC. The lower performance of RadViz can be due to the effect illustrated in Fig. 3a, where some points may be misinterpreted as outliers, and to the restriction that forces all mapped points to lie inside the convex hull of the anchor points (where distances are bounded). It is also apparent that the probabilities decrease as the number of variables increases, since the simultaneous effect of all the variables can hide the outlier. Lastly, for low likelihood outliers the probabilities are low and very similar for the three approaches. Thus, the radial projection methods are not mapping the outlier far enough from the rest of the points. These examples illustrate the importance of selecting multiple views (i.e., configurations of axis vectors or anchor points) when performing outlier detection with (radial) projection methods.

5 Linear Mappings

Another benefit of SC over RadViz is the possibility to select configurations of axis vectors in order to reproduce any linear mapping. This has practical benefits since it enables analysts to automatically generate useful plots, and visualize the configurations of axis vectors that lead to them. For example, the axis vectors can be chosen in order to reproduce popular linear transformations such as PCA, LDA, biplots, and so forth. These have been used successfully in order to perform or optimize tasks such as distance preservation, class or cluster separation, correlation approximation, etc. The following subsections show

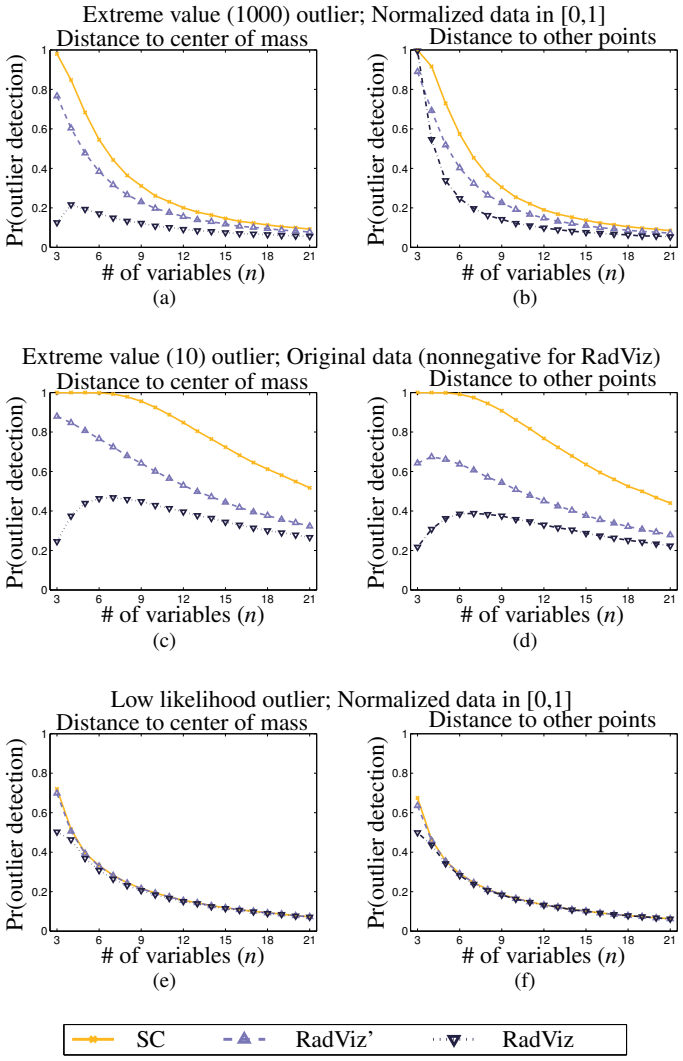


Fig. 8. Average outlier detection probabilities. In (a-d) the results involve detecting the outlier with an extreme value, while in (e) and (f) the task consist of finding the sample with low likelihood. The search for an outlier can involve calculating the distance to the center of mass (mean) of the plotted points, or to every other point in order to compute the (maximum) distance to the nearest neighbor. In (a) and (b) the extreme value of a single entry is 1000, while the data is later preprocessed to lie in the [0,1] interval. With SC it is easier to detect the outlier, but the performance drops quickly for both radial methods. In (c) and (d) we use the original (standardized) data, and the incorporated extreme value is 10. For RadViz we additionally shifted the data (by subtracting the minimum attribute values) in order to make them nonnegative. In this case the performance of the methods improves considerably, where the average probabilities are again larger for SC. Finally, in (e) and (f) we also preprocessed the data to lie in the [0,1] interval (we obtained similar results with the original data). The performance is poor, and almost identical, for both methods. The results correspond to basic SC plots, Radviz, and the variant in [28] for outlier detection, denoted as RadViz'.

several uses of this idea. In contrast, the inherent nonlinear transformation in RadViz impedes reproducing these useful plots.

5.1 Classification and feature selection

In practice it can be difficult to manually select a set of axis vectors in SC, or anchor points in RadViz, in order to separate classes in the observable display. A common approach in the literature consists of building regular configurations, and permuting the attributes in order

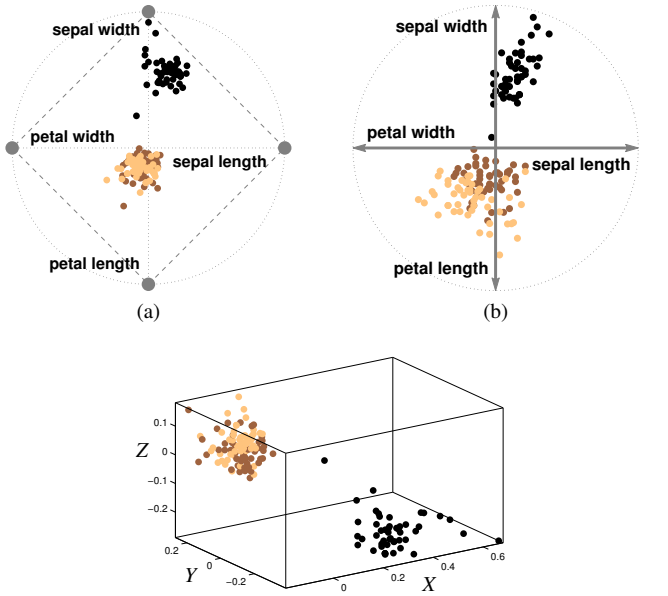


Fig. 9. Plots for a regular arrangement of the 4 variables of the Iris data set in RadViz (a), and SC (b). The lighter classes appear highly overlapped, especially in RadViz, since the nonlinear step maps them onto the same region of the 3-dimensional simplex ($\mathbf{1}^{\mathrm{T}}\mathbf{x} = 1$). Thus, there does not exist a configuration of anchor points that will separate those classes in RadViz.

to obtain different views of the data (see [3, 38]). However, the ordering of variables in visualizations is complex and time-consuming, since the problems to be solved are NP-hard, and therefore require heuristics, even for a moderate number of attributes (see [2, 5]). Moreover, regular configurations are often insufficient in order to obtain a satisfactory separation of classes on the plots. Figure 9 shows RadViz (a) and SC (b) plots for regular arrangements of the four variables (and 150 samples) of the Iris data set included in the UCI Machine Learning Repository [8]. The SC plot separates the lighter colored classes better than the analogous version in RadViz, but there nevertheless exists a clear overlap between them. Other orderings do not separate the classes either (see the supplemental material). In this example the high overlap in RadViz is due to its nonlinear step, which maps the lighter colored classes onto the same region of the unit 3-dimensional simplex, as shown in (c). Thus, no matter which configuration of anchor points we choose, those classes will always be overlapping in a RadViz plot.

However, when working with SC, analysts can take advantage of the numerous dimensionality reduction methods that generate linear mappings in order to separate a set of classes optimally, according to some criterion. The matrices associated with those mappings can be computed automatically, and specify the set of axis vectors to be used.

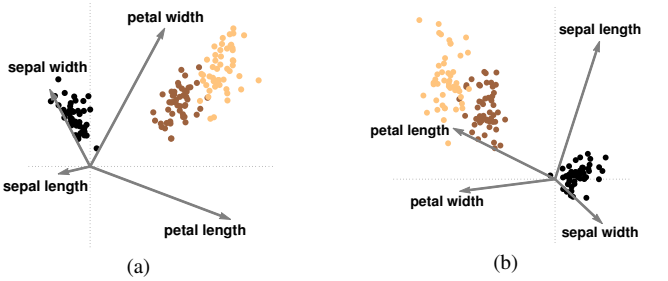


Fig. 10. Optimal configurations in SC for separating the classes of the Iris data set. In (a) the choice of axis vector produces the LDA plot, while in (b) the plot corresponds to the LMNN mapping.

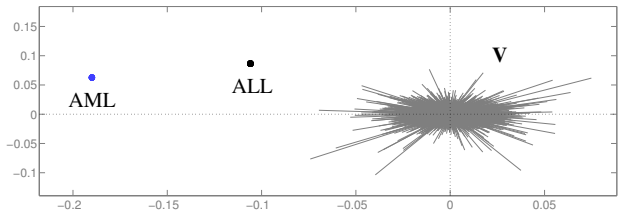


Fig. 11. Automatic configuration of 7129 genes (the gray lines represent the axis vectors) in a SC plot related to LDA. In the example, the data samples of two classes of leukemia (11 cases of AML and 27 of ALL) are clearly separated, and concentrated around two different points.

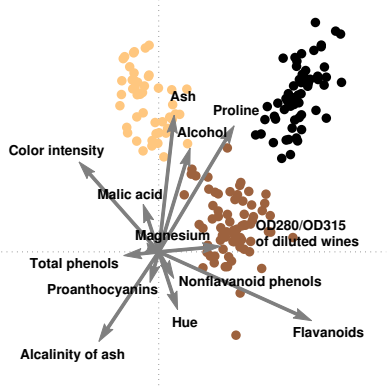
In SC it is straightforward to obtain these axis vectors. Consider some linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ defined by some known matrix $\mathbf{A}$, i.e., $\mathbf{p} = \mathbf{A}\mathbf{x}$. Due to (1) we can simply set $\mathbf{V} = \mathbf{A}^{\mathrm{T}}$ in order to reproduce the mapping. Thus, the axis vectors are simply the columns of $\mathbf{A}$. Figure 10 shows the configurations of axis vectors that generate (a) the LDA plot, and (b) the mapping associated with the large margin nearest neighbor (LMNN) [43] method, for the Iris data set. Both plots separate the lighter classes considerably better than regular configurations (see Fig. 9).

In Fig. 11 we applied LDA to the Gene expression data set used in [11], and analyzed in [13] with RadViz, in order to separate two classes of acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL). The resulting axis vectors ($\mathbf{V}$) are obtained automatically through the well-known statistical procedure, and the data is clearly separated on the plot, where the classes are mapped around two different points.

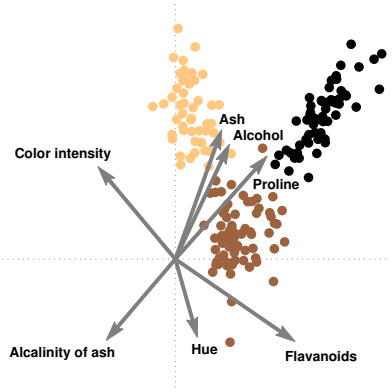
In certain tasks analysts may decide to simplify a visualization by discarding the variables that do not have a relevant contribution to it. In SC short axis vectors constitute possible candidates to be eliminated, since the effect of a variable in a plot depends on the length of its axis vector. For instance, when the set of axis vectors is obtained through the linear mapping corresponding to a method used for classification, it is possible to perform visual feature selection by discarding variables with short vectors, since these could be the least discriminative. Figure 12a shows the configuration of axis vectors associated with the LDA plot for the Wine data set in [8], which contains 13 variables and 178 samples. Since the three classes of wine are separated well on the plot, it may be possible to reduce the set of variables while still avoiding class overlaps. In (b) we have discarded six of the original variables (and reproduced the LDA plot for the remaining seven), which had the shorter axis vectors in (a). Note, however, that if there exist overlaps between classes, short vectors could play a relevant role in the plot. This occurs in Fig. 10a with the variable "sepal length" (note that its axis vector is oriented in the main direction that separates the lighter classes), or "sepal width" in (b).

Another approach designed for RadViz is the "class discrimination layout" (CDL) algorithm [26], which groups similar variables into sectors according to the t-statistic. Other procedures use quality measures that indicate class (or cluster) separation in order to select the locations or orderings of the anchor points. For instance, the approach described in [1] uses an image processing algorithm in order to separate classes. In Fig. 13 we show the result of applying these methods to the Wine data set (the anchor ordering in (b) is described in [1]). In order to evaluate class separation we have followed the approach in [21], which consists of computing the leave-one-out classification performance of a K-nearest neighbors classifier, for $K = 5$ (this choice generally provided optimal scores). The error rate for the SC plot related to LDA, shown in Fig. 12a, is not only much lower, but the axis vectors also reveal the importance of the variables in the plot.

Lastly, Fig. 14 shows another classification example taken from [1], which uses the Olives [45] data set containing 572 8-dimensional samples that are categorized in 9 classes. The error rate of a K-nearest neighbor classifier (for $K = 5$), applied to the optimal sector permutation of the CDL algorithm, shown in (a), is 12.41%. Note that since the data set contains 9 variables, there are $8!/2$ possible sector permutations. Since our implementation in MATLAB® takes about 4 seconds to evaluate the classification performance of one permutation (measured on an PC with an Intel® Core™ i7-4712HQ CPU, 2.30 GHz, and 16GB of RAM), the search for the optimal permutation took about 22.4 hours. In addition, the error rate for the class separation algorithm in [1], in (b), is 17.13% (zoomed-in figures are available in

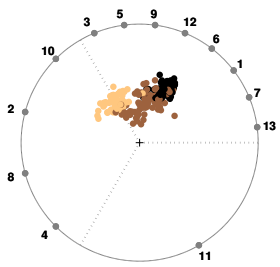


(a)

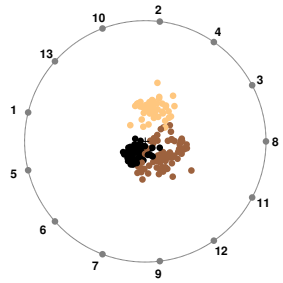


(b)

Fig. 12. Visual feature selection with SC. In (a) the configuration of axis vectors produces the LDA plot for the Wine data set. The least discriminative variables tend to have shorter vectors, and represent candidates to be discarded when performing feature selection. In (b) a reduced set of seven variables also separates the data well.



(a)



(b)

Fig. 13. RadViz class separation algorithms applied to the Wine data set. The error rate of a K-nearest neighbor classifier (for $K = 5$), applied to the optimal sector permutation of the CDL algorithm shown in (a), is 13.46%. The analogous error rate for the class separation algorithm in [1], shown in (b), is 4.49%. However, the error rate for the LDA plot in Fig. 12a is reduced to 0.56%.

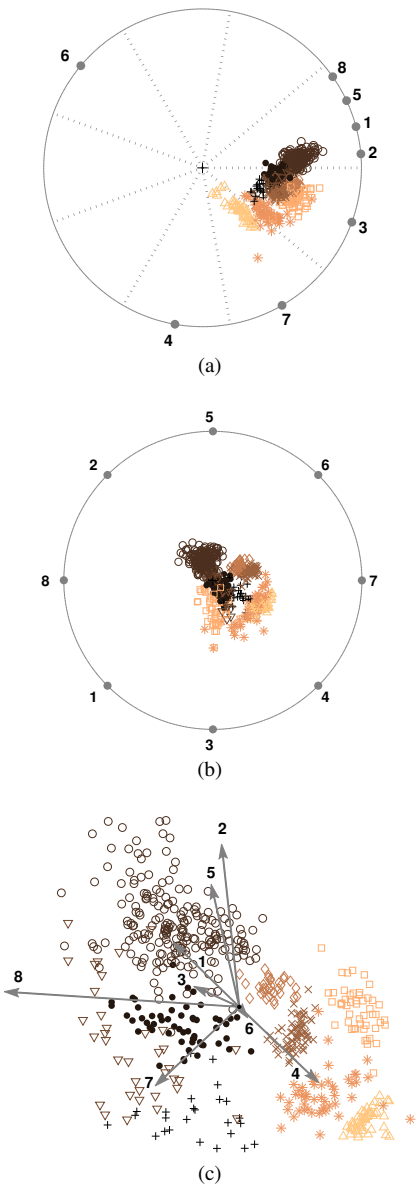


Fig. 14. Class separation algorithms applied to the Olives data set. The error rate of a K-nearest neighbor classifier (for $K = 5$), applied to the optimal sector permutation of the CDL algorithm for RadViz, shown in (a), is 12.41%. The analogous error rate for the class separation algorithm for RadViz in [1], shown in (b), is 17.13%. However, the combination of NCA and LMNN for SC, shown in (c), can decrease the error rate to 6.29%.

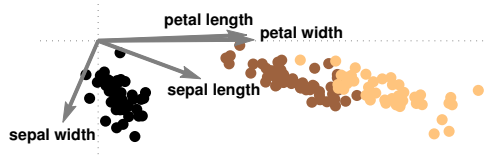


Fig. 15. Configuration of axis vectors in SC (for the Iris data set) that allows to estimate correlation coefficients as cosines of the angles between axis vectors.

the supplemental material).

Finally, an interesting property about linear maps is that they can be used consecutively. In Fig. 14c we have used neighborhood component analysis (NCA) [10] to map points onto a 2-dimensional space, generating a 2×8 matrix $\mathbf{A}$. Afterwards we have applied LMNN [43] on the corresponding mapping that provides a 2×2 matrix $\mathbf{B}$. Thus, the columns of matrix $\mathbf{BA}$ can be used to construct a SC plot. This approach can reduce the error considerably. In 20 random trials the average error rate was 7.57%, where the minimum was 6.29%. The combination of both approaches using [24], together with the classification performance evaluation, took 4.14 seconds on average. Therefore, this approach is considerably more efficient than searching for permutations.

Lastly, note that other nonlinear dimensionality reduction methods could separate the data even better. However, the linear methods produce matrices from which we can extract the SC axis vectors, which can be useful for understanding the role of the initial variables in the final visualization.

5.2 Biplots

Linear maps are useful for other analysis tasks. We conclude this section with an example involving statistical biplots. Note that an idea often used in visualizations based on radial axes consists of arranging similar attributes nearby [2, 3]. For example, it is possible to select the axis vectors in SC, or the anchor points in RadViz, in order to approximate correlation coefficients between the data variables. This can be carried out with the aid of classical statistical biplots, which build linear mappings that depict representations of both data samples and variables in a single graphic. In particular, they factor the best rank-2 approximation of the data (according to the Frobenius matrix norm) into two matrices that represent the points and variables. Let $\mathbf{USW}^{\mathrm{T}}$ be the singular value decomposition of the $N \times n$ data matrix, where N is the cardinality of the data set. By choosing $\mathbf{V}$ as the first two columns of $\mathbf{SW}^{\mathrm{T}}$, the resulting SC plot can be used to estimate correlation coefficients between variables, as the cosine between their respective axis vectors. Figure 15 shows this plot for the Iris data set, while Tab. 1 shows the different correlations and cosine values, which are similar.

6 Data Attribute Estimation

The mappings defined by SC and RadViz are not one-to-one since many high-dimensional samples can be projected onto the same low-dimensional point. Thus, when the samples are represented as dots, information is inevitably lost due to the dimensionality reduction process. In this regard, another benefit of SC over RadViz is related to the ability to recover original high-dimensional attribute values more accurately, simply from the visual elements in the plots. In particular, the axes in SC can be labeled, similarly to statistical biplots, in order to allow users to estimate values by projecting embedded points orthogonally onto them. It can be shown that, in order to obtain optimal estimates using this approach, the data should be centered, and matrix $\mathbf{V}$ should be orthogonal [32]. Figure 16 shows an example using four standardized variables of the US breakfast cereal data set used in [44] (it contains 77 samples but we discarded the three that have missing values). In the example, the configuration of axis vectors produces the PCA plot for the four attributes, where unhealthy cereals are represented by points on the right side of the plot, and healthy ones on the left. The caloric content for a particular cereal can be estimated by

Table 1. Correlation coefficients (r) and cosines between the axis vectors in Fig. 15. Note the similarity between the values.

Variable i	Variable j	r	$\cos(\mathbf{v}_i, \mathbf{v}_j)$
sepal length	sepal width	-0.1094	-0.0541
sepal length	petal length	0.8718	0.9237
sepal length	petal width	0.8180	0.9357
sepal width	petal length	-0.4205	-0.4324
sepal width	petal width	-0.3565	-0.4029
petal length	petal width	0.9628	0.9995

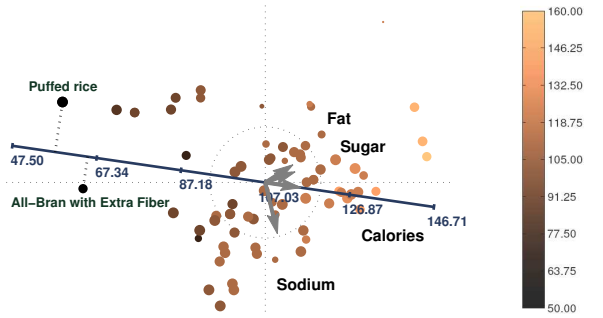


Fig. 16. Data attribute estimation in SC by projecting points orthogonally onto labeled axes. In the example, the caloric content of breakfast cereals can be estimated through projections onto the labeled axis. Additionally, the size of the points is related to the quality of the estimates (samples for which the estimates are accurate appear as larger dots). The color bar is associated with caloric content.

projecting its corresponding point onto the labeled axis. Lastly, it is possible to show the quality of the estimates by means of the size of the points.

The previous approach only requires modifying the axis vectors and centering the data. In contrast, the normalization step in RadViz transforms the data so that the sum of its elements is 1, which obviously eliminates information about the original attribute values. Thus, regardless of the placement of the anchor points and data translation, in RadViz users can try to compare whether an attribute value is greater for a sample than for another, but cannot estimate exact original values.

Lastly, for any dimensionality reduction method one way to visualize the exact attribute values consists of representing the mapped samples as glyphs or similar visual structures (see [40]) instead of points. However, this can only be carried out effectively for a low number of variables and data samples, since otherwise the plots would exhibit a considerable overlap.

7 Arbitrary Layouts

In SC, users can arrange the axis vectors interactively in arbitrary layouts in order to search for data with particular characteristics, cluster structure, or outliers (see [20, 44]). However, the anchor points in RadViz are usually arranged so that all of them form part of their convex hull. Although this is not strictly necessary, it is appropriate in order to analyze sparse data effectively. In addition, small convex hulls may be problematic, since the mapped points are always located inside them. Thus, we consider SC to be more flexible regarding user interaction and the number of layouts (i.e., views of the data) for the variables on the plots.

Figure 17 shows RadViz visualizations for the data set used in the SC plot in Fig. 16, where all of the axis vectors point to the same half-space (in part, because all of the variables are positively correlated, which occurs in many data sets). Note that it is useful to arrange the variables in that manner in SC in order to characterize healthy vs. unhealthy cereals. However, it is difficult to accomplish this effectively in RadViz with those four variables since points are pulled towards vertices of the convex hull, which cannot all have similar directions. Thus, it is not possible to position the anchors in order to detect data samples with large (or small) values for all attributes. In (a) the anchor points $\mathbf{v}_i$ are the same as in the SC plot. The resulting RadViz visualization is considerably different. On the one hand, the convex hull is particularly small. On the other hand, the ordering of the caloric content, which was apparent in the SC plot, is now degraded. In (b) we have updated the anchor points in order to form a regular layout, while preserving their relative ordering on the convex hull. In spite of this change, the mapped data points remain clearly unordered with respect to their caloric content.

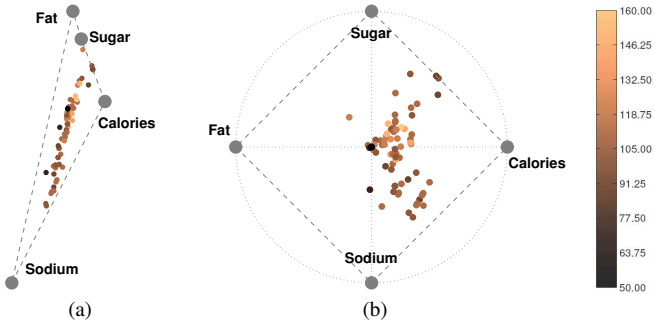


Fig. 17. RadViz applied to the data set used in Fig. 16. In (a) the anchor points are the same as in the SC plot. Their convex hull is small and the ordering of the caloric content is degraded. In (b) the related regular layout does not solve the problem.

8 Conclusions

This paper has compared SC and RadViz focusing on the effect of the latter's extra nonlinear normalization step, which is the main difference between them. We conclude that RadViz is especially advantageous when the data is sparse, as in [13, 37]. However, we have provided new results showing that its nonlinearity may hamper several other exploratory analysis tasks.

9 Discussion

Regarding the development of algorithms, many designed for RadViz can also be applied to SC. For instance, quality measures have been used to accomplish tasks such as class separation in RadViz [1]. Since these measures are general they can be applied to SC as well (and to other dimensionality reduction methods). Therefore, when developing algorithms for RadViz, we recommend considering whether they would also be appropriate for SC.

Regarding performance, SC and RadViz are equivalent since both compute the mapping of a data sample in $\Theta(n)$ time (necessary to calculate the product of a $2 \times n$ matrix times an n-dimensional vector). Radviz also requires calculating the sum of the attribute values of the samples, but this does not increase the computational complexity since it is also carried out in $\Theta(n)$ time. Thus, for a data set of cardinality N, both methods generate plots in $\Theta(Nn)$ time, and scale well even for very large data sets.

A common limitation of both methods is the difficulty to recover original data attributes, which can be mitigated in SC (see Sec. 6). Additionally, while modern dimensionality reduction methods have been designed to capture nonlinear structure of manifolds in the data, the methods under study are not capable of carrying out this task. On the one hand, SC is a linear method. On the other, RadViz defines a general nonlinear mapping that does not consider the shape or distribution of the data (i.e., the relationships between the samples). Thus, it is also incapable of revealing structure of manifolds in the data.

Lastly, we have performed an objective comparison of RadViz and SC based on theoretical analyses and experiments in which we have obtained results automatically, e.g., by modeling user behavior (see Sec. 4), or through computational procedures (see Sec. 5). In this regard, we are planning on performing user tests as a future work; for instance, in order to analyze interaction, and the ability to detect certain patterns, distributions, or correlations (see [16, 22]).

Acknowledgments

This work was supported by projects TIN2011-29542-C02-01 of the Spanish Ministry of Science and Innovation, the Human Brain Project, and the Cajal Blue Brain Project.

References

[1] G. Albuquerque, M. Eisemann, D. J. Lehmann, H. Theisel, and M. Magnor. Improving the visual analysis of high-dimensional datasets using quality-measures. In *IEEE Symposium on Visual Analytics Science and Technology (Proceedings IEEE VAST)*, 2010.

[2] M. Ankerst, S. Berchtold, and D. A. Keim. Similarity clustering of dimensions for an enhanced visualization of multidimensional data. In *Proceedings of the 1998 IEEE Symposium on Information Visualization*, INFOVIS '98, pages 52–60, Washington, DC, USA, 1998. IEEE Computer Society.

[3] A. O. Artero, M. F. de Oliveira, and H. Levkowitz. Enhanced high dimensional data visualization through dimension reduction and attribute arrangement. In *Proceedings of the International Conference on Information Visualization*, pages 707–712, Washington, DC, USA, 2006. IEEE Computer Society.

[4] E. Bertini, L. Dell' Aquila, and G. Santucci. Springview: Cooperation of radviz and parallel coordinates for view optimization and clutter reduction. In *Proceedings of the Coordinated and Multiple Views in Exploratory Visualization*, CMV '05, pages 22–29, Washington, DC, USA, 2005. IEEE Computer Society.

[5] L. D. Caro, V. Frias-Martinez, and E. Frias-Martinez. Analyzing the role of dimension arrangement for data visualization in radviz. In *Proceedings of the 14th Pacific-Asia conference on Advances in Knowledge Discovery and Data Mining - Volume Part II*, PAKDD'10, pages 125–132, 2010.

[6] K. Chen and L. Liu. iVIBRATE: Interactive visualization-based framework for clustering large datasets. *ACM Trans. Inf. Syst.*, 24(2):245–294, Apr. 2006.

[7] K. M. Daniels, G. G. Grinstein, A. Russell, and M. Glidden. Properties of normalized radial visualizations. *Information Visualization*, 11(4):273–300, 2012.

[8] A. Frank and A. Asuncion. UCI machine learning repository, 2010.

[9] K. R. Gabriel. The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58(3):453–467, Dec 1971.

[10] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov. Neighborhood component analysis. In *Advances in Neural Information Processing Systems 17*, pages 513–520, 2005.

[11] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286(5439):531–537, Oct 1999.

[12] G. G. Grinstein, P. E. Hoffman, and R. M. Pickett. Benchmark development for the evaluation of visualization for data mining. In U. Fayyad, G. G. Grinstein, and A. Wierse, editors, *Information Visualization in Data Mining and Knowledge Discovery*, pages 129–176. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2002.

[13] G. G. Grinstein, C. B. Jessee, P. E. Hoffman, P. J. O'Neil, and A. G. Gee. High-dimensional visualization support for data mining gene expression data. In E. V. Grigorenko, editor, *DNA Arrays: Technologies and Experimental Strategies*, chapter 6, pages 86–131. CRC Press LLC, Boca Raton, Florida, 2001.

[14] P. Hoffman, G. Grinstein, K. Marx, I. Grosse, and E. Stanley. DNA visual and analytic data mining. In *Proceedings of the 8th conference on Visualization '97*, VIS '97, pages 437–441, Los Alamitos, CA, USA, 1997. IEEE Computer Society Press.

[15] P. E. Hoffman. *Table Visualizations, a Formal Model and its applications*. PhD thesis, University of Massachusetts Lovell, 1999.

[16] P. Joia, F. Petronetto, and L. G. Nonato. Uncovering representative groups in multidimensional projections. *Computer Graphics Forum (Proc. EuroVis)*, 34(3): to appear, 2015.

[17] I. T. Jolliffe. *Principal component analysis*. Springer series in statistics. Springer-Verlag, 2010.

[18] M. C. Jones and R. Sibson. What is projection pursuit? *Journal of the Royal Statistical Society. Series A (General)*, 150(1):1–37, 1987.

[19] E. Kandogan. Star coordinates: A multi-dimensional visualization technique with uniform treatment of dimensions. In *In Proceedings of the IEEE Information Visualization Symposium, Late Breaking Hot Topics*, pages 9–12, 2000.

[20] E. Kandogan. 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*, KDD'01, pages 107–116, New York, NY, USA, 2001. ACM.

[21] G. Leban, B. Zupan, G. Vidmar, and I. Bratko. VizRank: Data Visualization Guided by Machine Learning. *Data Mining and Knowledge Discovery*, 13(2):119–136, Sept. 2006.

[22] D. J. Lehmann, F. Kemmler, T. Zhyhalava, M. Kirschke, and H. Theisel. Visualnostics: Visual guidance pictograms for analyzing projections of high-dimensional data. *Computer Graphics Forum (Proc. EuroVis)*, 34(3): to appear, 2015.

[23] D. J. Lehmann and H. Theisel. Orthographic star coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2615–2624, December 2013.

[24] L. v. Maaten. Matlab toolbox for dimensionality reduction, 2015.

[25] L. v. Maaten, E. Postma, and J. v. Herik. Dimensionality reduction: A comparative review. Technical Report TiCC-TR 2009-005, Tilburg University, 2009.

[26] J. F. McCarthy, K. A. Marx, P. E. Hoffman, A. G. Gee, P. O'Neil, M. L. Ujwal, and J. Hotchkiss. Applications of machine learning and high-dimensional visualization in cancer detection, diagnosis, and management. *Annals of the New York Academy of Sciences*, 1020:239–262, May 2004.

[27] G. J. McLachlan. *Discriminant analysis and statistical pattern recognition*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley-Interscience, 2004.

[28] L. Nováková. *Visualization data for Data Mining*. PhD thesis, Czech Technical University in Prague, 2009.

[29] R. M. Pillat and C. M. Freitas. Coordinating views in the infovis toolkit. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI'06, pages 496–499, New York, NY, USA, 2006. ACM.

[30] R. M. Pillat, E. R. A. Valiati, and C. M. Freitas. Experimental study on evaluation of multidimensional information visualization techniques. In *Proceedings of the 2005 Latin American Conference on Human-computer Interaction*, CLIHC'05, pages 20–30, New York, NY, USA, 2005. ACM.

[31] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[32] M. Rubio-Sánchez and A. Sanchez. Axis calibration for improving data attribute estimation in star coordinates plots. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2013–2022, December 2014.

[33] A. Russell, K. Daniels, and G. Grinstein. Voronoi diagram based dimensional anchor assessment for radial visualizations. In *Proceedings of the 2012 16th International Conference on Information Visualisation*, IV'12, pages 229–233, Washington, DC, USA, 2012. IEEE Computer Society.

[34] A. Russell, R. Marceau, F. Kamayou, K. Daniels, and G. Grinstein. Clustered data separation via barycentric radial visualization. In *Proceedings of the 2014 International Conference on Modeling, Simulation and Visualization Methods (MSV)*, pages 101–107, 2014.

[35] J. S. Shaik and M. Yeasin. Visualization of high dimensional data using an automated 3d star coordinate system. In *International joint conference on neural networks*, pages 1339–1346, 2006.

[36] J. Sharko. *Radviz Extensions with Applications*. PhD thesis, University of Massachusetts Lowell, 2009.

[37] J. Sharko, G. Grinstein, and K. A. Marx. Vectorized radviz and its application to multiple cluster datasets. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1444–1427, Nov. 2008.

[38] Y. Sun, J. Yuan, Y. Hu, and W. Xiao. An improved multivariate data visualization technique. In *International Conference on Information and Automation, ICIA'08.*, pages 1525–1530, june 2008.

[39] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[40] H. Theisel and M. Kreuseler. An enhanced spring model for information visualization. *Computer Graphics Forum*, 17(3):335–344, 1998.

[41] C.-Y. Tsai and C.-C. Chiu. A clustering-oriented star coordinate translation method for reliable clustering parameterization. In *Proceedings of the 12th Pacific-Asia conference on Advances in knowledge discovery and data mining*, PAKDD'08, pages 749–758, Berlin, Heidelberg, 2008. Springer-Verlag.

[42] E. R. Valiati, C. M. Freitas, and M. S. Pimenta. Using multi-dimensional in-depth long-term case studies for information visualization evaluation. In *Proceedings of the 2008 Workshop on BEyond Time and Errors: Novel evaLuation Methods for Information Visualization*, BELIV'08, pages 1–7, New York, NY, USA, 2008. ACM.

[43] K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 10:207–244, June 2009.

[44] J. S. Yi, R. Melton, J. Stasko, and J. A. Jacko. Dust & magnet: multivariate information visualization using a magnet metaphor. *Information Visualization*, 4(4):239–256, 2005.

[45] J. Zupan, M. Novic, X. Li, and J. Gasteiger. Classification of multicomponent analytical data of olive oils using different neural networks. *Analytica Chimica Acta*, 292(3):219–234, July 1994.