Axis Calibration for Improving Data Attribute Estimation in Star Coordinates Plots

Manuel Rubio-Sánchez and Alberto Sanchez

Abstract—Star coordinates is a well-known multivariate visualization method that produces linear dimensionality reduction mappings through a set of radial axes defined by vectors in an observable space. One of its main drawbacks concerns the difficulty to recover attributes of data samples accurately, which typically lie in the [0,1] interval, given the locations of the low-dimensional embeddings and the vectors. In this paper we show that centering the data can considerably increase attribute estimation accuracy, where data values can be read off approximately by projecting embedded points onto calibrated (i.e., labeled) axes, similarly to classical statistical biplots. In addition, this idea can be coupled with a recently developed orthonormalization process on the axis vectors that prevents unnecessary distortions. We demonstrate that the combination of both approaches not only enhances the estimates, but also provides more faithful representations of the data.

Index Terms—Star Coordinates, RadViz, Biplots, Axis calibration, Attribute value estimation, Data centering, Orthographic projection

◆

Exploratory data analysis is an approach aimed at obtaining overviews of data sets, and consists of an integral component of a data mining process. Many of the techniques generate visual representations of data sets, where users are integrated into the analysis process with the goal of combining their perceptual capabilities, flexibility, or domain knowledge with the computational power of today's computers.

Multivariate visualization techniques can be categorized according to different criteria, including data type, the graphical objects and layouts that compose the plots, or forms of interaction [16]. Some are capable of displaying high-dimensional data without any loss of information (e.g., scatterplot matrices, parallel coordinates [12, 22, 11], or the table lens [19]), which show exact attribute values of data samples directly on the plots. In contrast, in this paper we focus on methods based on radial axes (see [5, 4]) that generate transformations of data onto an observable display, where information is inevitably lost in the dimensionality reduction process. For these methods users generally cannot determine precise values of data samples by just observing their low-dimensional representations. Thus, users can only recover original data values visually to a certain extent. In this paper we refer to this process as attribute (or data value) estimation. Although these methods are useful for exploratory purposes in the overview phase of the well-known information seeking mantra [21], the differences between the true and estimated data values may cause misinterpretations.

In this paper we study attribute estimation accuracy in Star Coordinates (SC) [14, 15, 20, 25, 23, 18], which is acknowledged to be one of the method's major drawbacks [5]. Our goal is to provide an in-depth analysis of several operations that allow to obtain more accurate estimates, and help users compare, group, examine, and understand the data quickly, without the need to inspect numerical values directly. In particular, we show that the accuracy of the estimates can be enhanced by reading off data values through projections of embedded points onto "calibrated" (i.e., labeled) axes, similarly to classical statistical biplots. When using this approach we prove that estimation accuracy can be increased considerably by centering the data. Thus, the usual preprocessing operation that normalizes the data to lie in the [0,1] interval is not appropriate for estimating attribute values through projections onto calibrated axes. Additionally, it has been shown recently that the axis vectors can be modified or restricted in order to

---

- *Manuel Rubio-Sánchez is with URJC. E-mail: manuel.rubio@urjc.es.*
- *Alberto Sanchez is with URJC. E-mail: alberto.sanchez@urjc.es.*

generate orthographic projections in SC plots [17], which also helps to reduce estimation errors, although at a lesser extent than the centering approach. We demonstrate the accuracy improvement through a comparative study that includes models of how users estimate data values in SC and RadViz [10, 3], which is another popular method based on radial axes. Finally, we show that the combination of the centering and the orthographic projections not only enhances the estimates in SC, but also provides more faithful representations of the data.

The rest of the paper is organized as follows. Section 1 reviews the related work. Section 2 describes the theory underlying the proposed approach, while Sec. 3 presents experimental results. Finally, Sec. 4 provides a discussion.

1 RELATED WORK

Star coordinates (see Fig. 1) is a simple, efficient, and well known interactive method for multivariate data visualization that can be used for exploratory purposes including cluster analysis, outlier and trend detection, or decision making tasks. The method generates linear dimensionality reduction mappings from an n-dimensional data space to 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. 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:

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

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.

Any dimensionality reduction process naturally entails a loss of information. Thus, in SC users will not be able to recover attributes of data samples perfectly, given their embedded points and the axis vectors. In addition, estimating attribute values is particularly difficult in SC [5]. When thinking of the linear combination, the estimates of attribute values are not independent of each other, since assigning some particular value to a variable affects the possible values of the estimates for the rest of the variables. In this regard, the estimation process can be understood graphically as building a "path" from the origin to a known $\mathbf{p}$, by concatenating scaled axis vectors of length $\hat{x}_i \cdot \|\mathbf{v}_i\|$, where $\hat{x}_i$ is the estimate of x_i, similarly to the SC mapping procedure, as shown in Fig. 1. Thus, fixing a particular estimate (which defines the length of a segment) affects the values of the rest. When estimating attributes through this approach users must therefore consider all of the
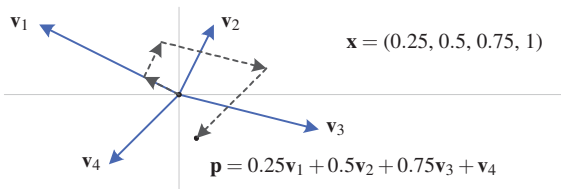
Fig. 1. Diagram of Star Coordinates' linear mapping. A 4-dimensional data sample $\mathbf{x} = (0.25, 0.5, 0.75, 1)$ gets mapped onto the 2-dimensional point $\mathbf{p}$ by adding scaled versions of the vectors $\mathbf{v}_i$, where the scaling factors are the values of $\mathbf{x}$ associated with each variable. The process can be understood graphically as building a "path" from the origin to $\mathbf{p}$.

possible estimates simultaneously, which is complicated, especially as the number of variables increases.

In order to mitigate this issue, we focus on optimizing the estimates when they correspond to orthogonal projections of embedded points onto axes, similarly to biplots [7, 9, 8]. These classical statistical visualizations essentially form scaled versions of the principal component analysis (PCA) plot, representing the best rank m approximation of the data in a least squares sense (see [13]). One of the most interesting properties of biplots is the ability to read off estimates of data attributes by projecting embedded points orthogonally onto adequately calibrated axes, as shown in Fig. 2. In biplots, the estimate $\hat{x}_i$ for the i-th attribute is defined as a dot product:

$$\hat{x}_i = \mathbf{v}_i^{\mathsf{T}}\mathbf{p} = \frac{\mathbf{v}_i^{\mathsf{T}}\mathbf{p}}{\|\mathbf{v}_i\|}\|\mathbf{v}_i\|.$$

Thus, the (orthogonal) scalar projections must be multiplied by the lengths of the axis vectors, where the distance between consecutive integers on the i-th axis is $1/\|\mathbf{v}_i\|$. Finally, the estimate $\hat{\mathbf{x}}$ of a data sample is therefore:

$$\hat{\mathbf{x}} = \mathbf{V}\mathbf{p}. \tag{2}$$

In Section 3.1 we describe a model of how users typically estimate values in SC.

Another well known technique related to SC is RadViz, where the vectors $\mathbf{v}_i$ define anchor points of springs in a physical model rather than axes (see [10, 3]). In particular, it generates nonlinear projections according to:

$$\mathbf{p} = \frac{\sum_{i=1}^{n} x_i \mathbf{v}_i}{\sum_{i=1}^{n} x_i} = \mathbf{V}^{\mathsf{T}}\left(\frac{\mathbf{x}}{\mathbf{1}^{\mathsf{T}}\mathbf{x}}\right), \tag{3}$$

where the data is normalized so that the range of each variable is the $[0, 1]$ interval. Nevertheless, (3) shows that RadViz is equivalent to SC if each data sample is also processed so that the sum of its elements is equal to 1. Note that this last normalization step produces nonlinear mappings of the data onto an observable display, since it applies a linear mapping (SC) to points projected nonlinearly onto the unit $(n − 1)$-simplex ($\mathbf{1}^{\mathsf{T}}\mathbf{x} = 1$).

In RadViz the anchor points are usually arranged in a regular pattern, although it is possible to update their location interactively. The embedded points $\mathbf{p}$ will lie inside the convex hull defined by the anchor points, since the former are convex combinations of the latter. In regular patterns the method has 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 i-th element is the only nonzero one are mapped to $\mathbf{v}_i$, (3) samples whose data attributes are all the same get mapped to the origin, and (4) when the embedding $\mathbf{p}$ is located on an edge of the convex hull between $\mathbf{v}_i$ and $\mathbf{v}_j$, then the attributes for the remaining variables will be 0. In Section 3.2 we describe a model of how users can estimate values in RadViz that is consistent with these properties.

2 CALIBRATION AND OPTIMAL ESTIMATES

The estimation error associated with approximating data attributes through projections onto calibrated axes as in biplots, through (2), can be defined as:

$$\varepsilon = \|\hat{\mathbf{x}} - \mathbf{x}\| = \|\mathbf{V}\mathbf{p} - \mathbf{x}\|,$$

where $\|\cdot\|$ is some vector norm (in this paper we use the Euclidean norm or its square). For SC, the estimation error is therefore:

$$\varepsilon = \|\mathbf{V}\mathbf{V}^{\mathsf{T}}\mathbf{x} - \mathbf{x}\|, \tag{4}$$

due to (1). The following subsections describe approaches that minimize this error.

2.1 Data centering

When using (4) the data preprocessing approach turns out to be fundamental in order to estimate values accurately across an entire data set. Firstly, note that, due to (2), the estimate $\hat{\mathbf{x}}$ of a data sample is a linear combination of the columns of $\mathbf{V}$. Therefore, it lies on the range or "column space" of $\mathbf{V}$, which is the subspace spanned by the columns of $\mathbf{V}$. In this paper it is a plane embedded in $\mathbb{R}^n$ since we only consider the case when $m = 2$. In the remainder of the paper we refer to this subspace as $\mathscr{R}(\mathbf{V})$, where $\mathscr{R}$ denotes range. Additionally, the distance between $\mathbf{x}$ and its estimate $\hat{\mathbf{x}}$ is precisely the estimation error. Thus, it is also related to the distance from $\mathbf{x}$ to $\mathscr{R}(\mathbf{V})$, where it is apparent that the data should lie close to that subspace in order to obtain accurate estimates (see Fig. 3). The following results show that, in order to minimize the sum of squared estimation errors for an entire data set, the data can be shifted so that its new mean lies in $\mathscr{R}(\mathbf{V})$, which may as well be the origin. Thus, centering the data will allow to obtain more accurate estimates in general.

Proposition 1. *Let $\mathbf{V}$ be a full column rank matrix whose rows contain the axis vectors in the SC model. Also, consider a data set of cardinality N, where $\mathbf{x}_i$ is the i-th data sample, and $\mathbf{p}_i$ its embedded point. In that case, the sum of squared estimation errors for all of the points in a new data set that has been shifted by some vector $\mathbf{t}$:*

$$\sum_{i=1}^{N} \|\mathbf{V}\mathbf{p}_i - (\mathbf{x}_i - \mathbf{t})\|^2, \tag{5}$$
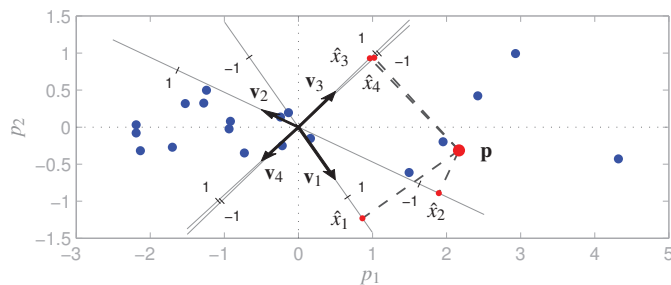

Fig. 2. Estimation of data values through projections in biplots. The estimate $\hat{\mathbf{x}} \in \mathbb{R}^4$ of a data sample $\mathbf{x} \in \mathbb{R}^4$ that has been embedded onto $\mathbf{p} \in \mathbb{R}^2$, is found by projecting $\mathbf{p}$ onto adequately calibrated axes associated with the variables, and defined by the axis vectors $\mathbf{v}_i$.
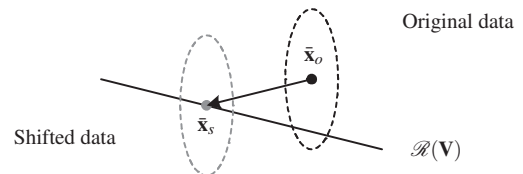

Fig. 3. Consider a particular configuration of axis vectors determined by $\mathbf{V}$. In that case the original data (with mean $\bar{\mathbf{x}}_o$) should be shifted so that its new mean ($\bar{\mathbf{x}}_s$) lies on $\mathscr{R}(\mathbf{V})$ in order to obtain more accurate estimates through projections, since estimation errors are related to distances from points to $\mathscr{R}(\mathbf{V})$. The shifting operation reduces these distances on average, and therefore the estimation errors.
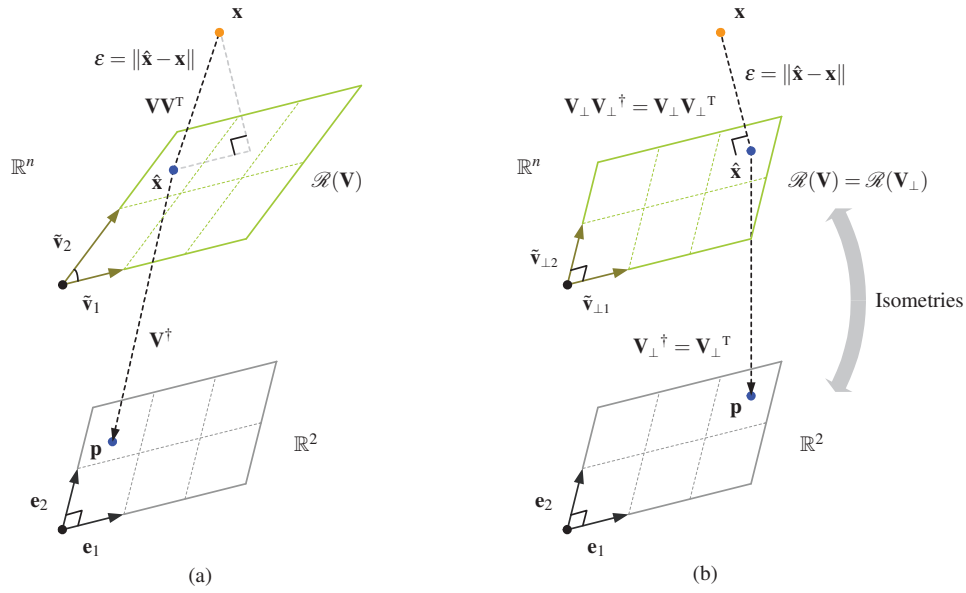
Fig. 4. Two-step interpretation of the linear mapping defined by SC. Data samples are first projected onto $\mathscr{R}(\mathbf{V})$ through $\mathbf{V}\mathbf{V}^{\mathrm{T}}$, and afterwards mapped onto the observable space ($\mathbb{R}^2$), through $\mathbf{V}^{\dagger}$. Roughly speaking, this last operation performs rotations and scalings in order to align $\tilde{\mathbf{v}}_i$ with $\mathbf{e}_i$. If $\tilde{\mathbf{v}}_1$ and $\tilde{\mathbf{v}}_2$ (the columns of $\mathbf{V}$) do not form an orthonormal set, as shown in (a), the first projection is oblique leading to larger estimation errors ($\varepsilon = \|\hat{\mathbf{x}} - \mathbf{x}\|$), while distances and angles in $\mathscr{R}(\mathbf{V})$ are not preserved in $\mathbb{R}^2$. In (b), $\tilde{\mathbf{v}}_{\perp 1}$ and $\tilde{\mathbf{v}}_{\perp 2}$ form an orthonormal set, where not only is the estimate $\hat{\mathbf{x}}$ optimal (since it is the orthogonal projection of $\mathbf{x}$ onto $\mathscr{R}(\mathbf{V})$), but distances and angles in $\mathscr{R}(\mathbf{V})$ are preserved in $\mathbb{R}^2$ due to the isometries between both spaces.

is minimized if $\mathbf{t} = \bar{\mathbf{x}} + \mathbf{z}$, where $\bar{\mathbf{x}}$ is the sample mean of the data, and $\mathbf{z}$ is any vector in the range of $\mathbf{V}$.

Proof. The optimal shift $\mathbf{t}$ of the data is found by minimizing (5). Since $\mathbf{p} = \mathbf{V}^{\mathrm{T}}\mathbf{x}$ in SC, it follows that $\mathbf{p}_i = \mathbf{V}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{t})$ for the i-th data point. Thus, the function to be minimized is:

$$f(\mathbf{t}) = \sum_{i=1}^{N} \|\mathbf{V}\mathbf{V}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{t}) - (\mathbf{x}_i - \mathbf{t})\|^2$$

$$= \sum_{i=1}^{N} \|(\mathbf{I} - \mathbf{V}\mathbf{V}^{\mathrm{T}})\mathbf{t} - (\mathbf{I} - \mathbf{V}\mathbf{V}^{\mathrm{T}})\mathbf{x}_i\|^2$$

$$= \sum_{i=1}^{N} \|\mathbf{B}\mathbf{t} - \mathbf{c}_i\|^2 = \sum_{i=1}^{N} \left(\mathbf{t}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{t} - 2\mathbf{t}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{c}_i + \mathbf{c}_i^{\mathrm{T}}\mathbf{c}_i\right)$$

$$= N\mathbf{t}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{t} - 2\mathbf{t}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\sum_{i=1}^{N} \mathbf{c}_i + \sum_{i=1}^{N} \mathbf{c}_i^{\mathrm{T}}\mathbf{c}_i,$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{B} = (\mathbf{I} - \mathbf{V}\mathbf{V}^{\mathrm{T}})$, and $\mathbf{c}_i = \mathbf{B}\mathbf{x}_i$. Taking derivatives with respect to $\mathbf{t}$ and setting them to $\mathbf{0}$:

$$\frac{\partial f}{\partial \mathbf{t}} = 2N\mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{t} - 2\mathbf{B}^{\mathrm{T}}\sum_{i=1}^{N} \mathbf{c}_i = \mathbf{0},$$

$$2N\mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{t} - 2\mathbf{B}^{\mathrm{T}}\mathbf{B}\sum_{i=1}^{N} \mathbf{x}_i = \mathbf{0},$$

$$\mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{t} - \frac{1}{N}\mathbf{B}^{\mathrm{T}}\mathbf{B}\sum_{i=1}^{N} \mathbf{x}_i = \mathbf{B}^{\mathrm{T}}\mathbf{B}(\mathbf{t} - \bar{\mathbf{x}}) = \mathbf{0}.$$

Thus, $\mathbf{t} = \bar{\mathbf{x}} + \mathbf{z}$, where $\mathbf{z} \in \mathscr{R}(\mathbf{V})$, since $\mathscr{R}(\mathbf{V}) = \mathscr{R}(\mathbf{V}\mathbf{V}^{\mathrm{T}}) = \mathscr{R}(\mathbf{I} - \mathbf{V}\mathbf{V}^{\mathrm{T}})^{\perp} = \mathscr{N}(\mathbf{I} - \mathbf{V}\mathbf{V}^{\mathrm{T}}) = \mathscr{N}(\mathbf{B})$, where $\mathscr{N}$ denotes nullspace. $\square$

Corollary 1. *In SC the data can be centered in order to minimize the sum of squared estimation errors for all of the points in a data set when $\mathbf{V}$ has full column rank.*

Proof. From Prop. 1 it follows that $\mathbf{t}$ can be $\bar{\mathbf{x}}$, since we can choose $\mathbf{z} = \mathbf{0}$. $\square$

2.2 Orthonormalization of the columns of $\mathbf{V}$

Recently, Lehmann and Theisel [17] proposed to restrict SC to orthographic projections in the context of interaction in order to improve the quality of the visualizations. The idea consists of replacing the columns of $\mathbf{V}$ by an orthonormal set of vectors that span the same subspace as the original ones. This generates a new set of axis vectors that produces a similar plot, where the estimates for each data sample are optimal, and which leads to more faithful representations of the data. In this section we review the approach but focusing on enhancing attribute estimation accuracy.

Firstly, note that the SC mapping (see (1)) can be decomposed into two separate linear transformations:

$$\mathbf{p} = \mathbf{V}^{\mathrm{T}}\mathbf{x} = [\mathbf{V}^{\dagger}][\mathbf{V}\mathbf{V}^{\mathrm{T}}]\mathbf{x} = \mathbf{V}^{\dagger}\hat{\mathbf{x}},$$

where † denotes the Moore-Penrose pseudoinverse (we assume that $\mathbf{V}$ has full column rank, where $\mathbf{V}^{\dagger} = (\mathbf{V}^{\mathrm{T}}\mathbf{V})^{-1}\mathbf{V}^{\mathrm{T}}$). Geometrically, the idea is illustrated in Fig. 4a. Thus, we can interpret that SC first projects data samples onto $\mathscr{R}(\mathbf{V})$, through $\mathbf{V}\mathbf{V}^{\mathrm{T}}$, where the resulting points constitute the n-dimensional estimates. In order to complete the process, the second mapping defined by $\mathbf{V}^{\dagger}$ simply rotates and scales the estimates in $\mathscr{R}(\mathbf{V})$ in order to represent them in the observable space ($\mathbb{R}^m$). In particular, the column vectors of $\mathbf{V}$, denoted as $\tilde{\mathbf{v}}_i$, for $i = 1, \ldots, m$, are transformed into the vectors of the standard basis ($\mathbf{e}_i$) of the observable space.

Since the final plot can be interpreted as a simple linear transformation of the information on $\mathscr{R}(\mathbf{V})$, we can substitute the column vectors of $\mathbf{V}$ by an orthonormal set that spans the same subspace, as shown in Fig. 4b, where $\mathbf{V}_{\perp}$ denotes this new orthogonal matrix of axis vectors (thus, $\mathbf{V}_{\perp}^{\mathrm{T}}\mathbf{V}_{\perp} = \mathbf{I}$), and where $\tilde{\mathbf{v}}_{\perp i}$, for $i = 1, \ldots, m$, are the corresponding column vectors. It is straightforward to compute $\mathbf{V}_{\perp}$, for example,
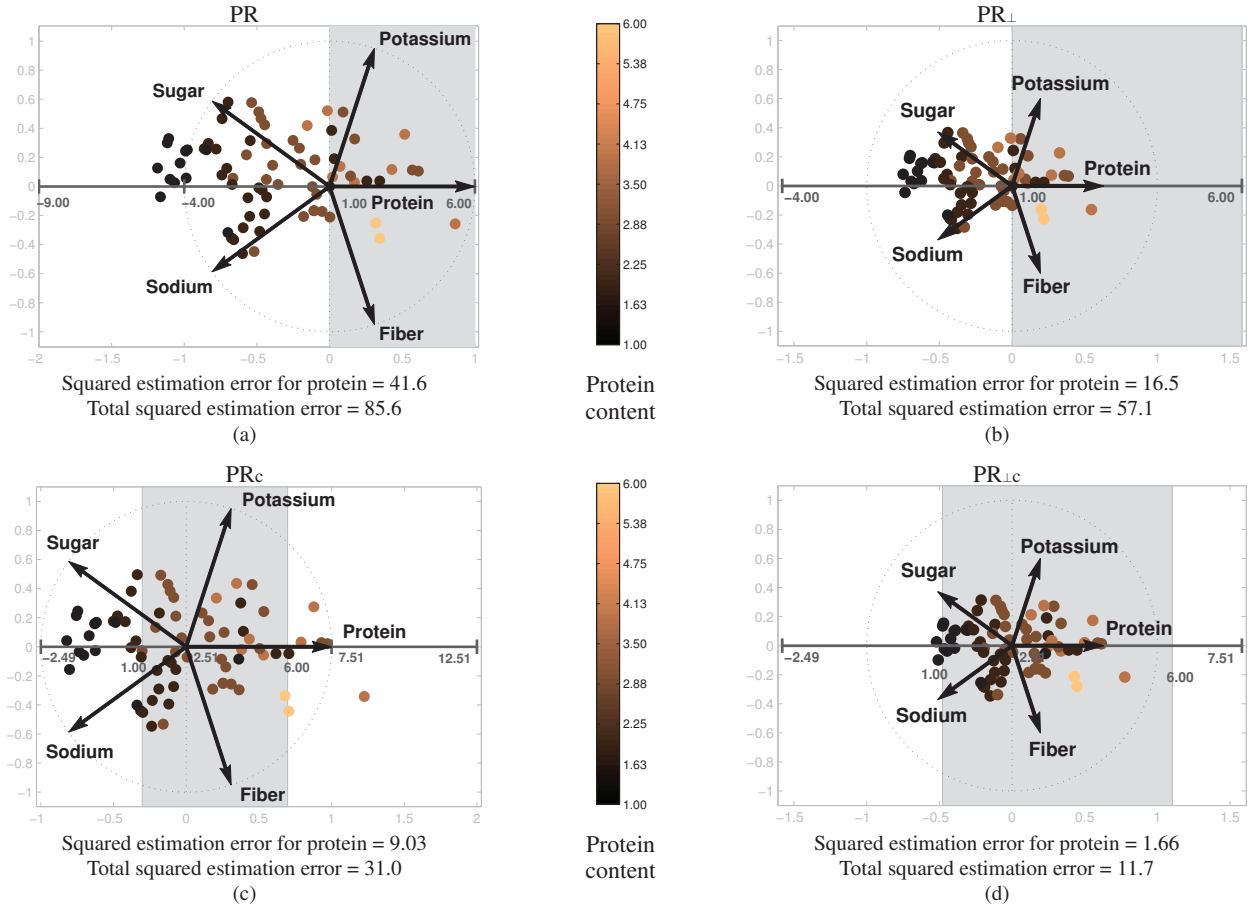
Fig. 5. Comparison of methods. The graphics show SC plots of a breakfast cereal data set containing five variables, where the axis associated with protein content has been labeled with original values that lie in the interval [1,6], shown as a gray area. The points have been colored according to protein content, where the color bar indicates the particular coding. Note that lighter points, with higher values, are located towards the right, in the direction of the axis vector. Additionally, since the estimates for protein are the (orthogonal) projections of the points onto its labeled axis, ideally they should be located inside the gray area. In (a) the plot uses data in [0,1] (not centered) and a configuration of axis vectors that does not generate orthographic projections, where the estimates are highly inaccurate. In (b) the columns of $\mathbf{V}$ form an orthonormal set, which has a positive effect on the accuracy. In (c) the vectors do not produce orthographic projections but the [0,1] data has been centered, which allows to reduce the squared estimation errors considerably. Finally, in (d) both approaches are combined to reduce the attribute estimation error even further, where almost every point lies in the [1,6] interval.

through the Gram-Schmidt orthonormalization procedure ([17] also proposes a gradient descent based algorithm which essentially minimizes $\|\mathbf{V}^{\mathsf{T}}\mathbf{V} - \mathbf{I}\|^2_{\text{Fro}}$, where Fro denotes the Frobenius norm). This defines an alternative configuration of axis vectors, but has several benefits that we describe below.

On the one hand, recall that the estimation errors are the distances from $\mathbf{x}$ to $\hat{\mathbf{x}}$. Thus, these are minimized for individual points if $\hat{\mathbf{x}}$ is the orthogonal projection of $\mathbf{x}$ onto $\mathscr{R}(\mathbf{V})$. In SC the optimal estimates are achieved if and only if $\mathbf{V}\mathbf{V}^{\mathsf{T}}$ is an orthogonal projection matrix, and this occurs if and only if the columns of $\mathscr{R}(\mathbf{V})$ form an orthonormal set of vectors (i.e., if and only if $\mathbf{V}^{\mathsf{T}}\mathbf{V} = \mathbf{I}$). If the column vectors of $\mathbf{V}$ do not form an orthonormal set the projections are oblique and can lead to large estimation errors.

On the other hand, when using $\mathbf{V}_\perp$ dot products, and therefore distances and angles, between points that lie on $\mathscr{R}(\mathbf{V})$ are preserved exactly on the low dimensional display, due to the following result.

Proposition 2. *The linear mapping from the observable inner product space $\mathbb{R}^m$ to $\mathscr{R}(\mathbf{V})$, defined by $\mathbf{V}$, as well as the transformation from $\mathscr{R}(\mathbf{V})$ to $\mathbb{R}^m$, defined by $\mathbf{V}^\dagger$, will be isometries if and only if the columns of $\mathbf{V}$ form a orthonormal set of vectors (i.e., if and only if $\mathbf{V}^{\mathsf{T}}\mathbf{V} = \mathbf{I}$).*

Proof. Consider two points $\mathbf{p}, \mathbf{q} \in \mathbb{R}^m$, and their images $\mathbf{V}\mathbf{p}, \mathbf{V}\mathbf{q} \in$ $\mathscr{R}(\mathbf{V})$, then:

$$\langle \mathbf{p}, \mathbf{q} \rangle = \mathbf{p}^{\mathsf{T}}\mathbf{q} = \mathbf{p}^{\mathsf{T}}\mathbf{V}^{\mathsf{T}}\mathbf{V}\mathbf{q} = \langle \mathbf{V}\mathbf{p}, \mathbf{V}\mathbf{q} \rangle \Leftrightarrow \mathbf{V}^{\mathsf{T}}\mathbf{V} = \mathbf{I},$$

where $\langle \cdot, \cdot \rangle$ denotes dot product. Similarly, the linear mapping from $\mathscr{R}(\mathbf{V})$ to $\mathbb{R}^m$ defined by $\mathbf{V}^\dagger$ will also be an isometry:

$$\langle \mathbf{V}\mathbf{p}, \mathbf{V}\mathbf{q} \rangle = \langle \mathbf{p}, \mathbf{q} \rangle = \langle \mathbf{V}^\dagger\mathbf{V}\mathbf{p}, \mathbf{V}^\dagger\mathbf{V}\mathbf{q} \rangle,$$

if and only if $\mathbf{V}$ defines an isometry from $\mathbb{R}^m$ to $\mathscr{R}(\mathbf{V})$, which happens if and only if $\mathbf{V}^{\mathsf{T}}\mathbf{V} = \mathbf{I}$, as proved above. $\qquad\square$

Finally, we provide a result that shows that the lengths of the axis vectors are at most 1 when the columns of $\mathbf{V}$ form an orthonormal set of vectors:

Proposition 3. *Let $\mathbf{V}$ be an $n \times m$ matrix, where $n \geq m$, whose columns form an orthonormal set of vectors. The Euclidean norm of any row of $\mathbf{V}$ will be at most 1.*

Proof. Firstly, the entries of $\mathbf{V}$ are constrained since the Euclidean norms of the columns are equal to 1. Additionally, since $\mathbf{V}$ has rank m, in order to find the largest squared values for the entries of some rows we can assume that all of the entries of $n - m$ rows of $\mathbf{V}$ are 0. In this optimal scenario, the remaining m rows form a matrix $\mathbf{R}$ which is orthogonal, where $\mathbf{R}^{\mathsf{T}}\mathbf{R} = \mathbf{R}\mathbf{R}^{\mathsf{T}} = \mathbf{I}$, which implies that the rows of $\mathbf{R}$ have unit Euclidean norm. $\qquad\square$

2.3 Benefits of axis calibration for data analysis

The new visualizations with centered data and $\mathbf{V}_\perp$ provide more faithful representations that can be useful in tasks where users need to approximate data values quickly. Figure 5 shows the benefit on estimation accuracy of combining both approaches, on configurations of five variables of the breakfast cereal data set available in lib.stat.cmu.edu/datasets/1993.expo/. For reference, we have labeled (calibrated) the axis associated with protein content with the original values (in grams), which lie in $[1,6]$, and have shaded the area corresponding to such interval. Thus, if an embedded point falls outside of the gray area its orthogonal projection onto the protein axis will result in a value that is outside of $[1,6]$. Lastly, we have colored the embedded samples according to their protein content (the color bars in the middle show the particular color coding).

In Fig. 5a the plot is generated with data normalized to lie in $[0,1]$, and is therefore not centered. Regarding the configuration of axis vectors, when they are regular, as in the graphics, they only provide orthographic projections when the lengths of the axis vectors are $\sqrt{2/n}$ (where n is the number of variables) [17]. In Fig. 5a their length is 1, and therefore $\mathbf{V}$ is not an orthogonal matrix. The estimates associated with the resulting plot are therefore highly inaccurate, where more than half of the points are not even in the gray area (note that some have large negative values). In Fig. 5b the lengths of the axis vectors are $\sqrt{2/5}$, which cause the columns of $\mathbf{V}$ to form an orthonormal set of vectors. This modifies the scaling on the axes, which has a positive effect on the estimation accuracy, even though the number of points that fall inside the gray area is the same as in Fig. 5a. In Fig. 5c the vectors have unit length but the data (previously in $[0,1]$) has been centered, which reduces the squared estimation error considerably. Finally, in Fig. 5d we have used this centered data and an orthogonal matrix, where most of the points fall inside the gray area, and provide the lowest squared estimation errors.

In addition, when users inspect data for decision making a commonly used approach consists of pointing vectors associated with desirable features in a similar direction, while pointing undesirable vari-
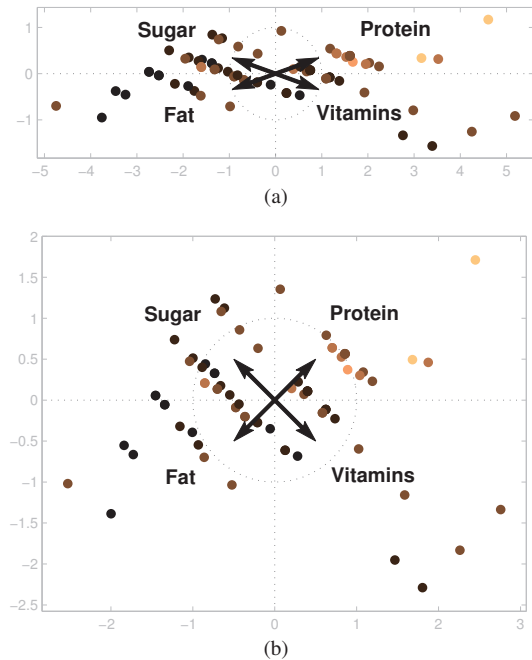

Fig. 6. Avoiding unnecessary distortions. In (a) the axis vectors are chosen in order to characterize health vs. unhealthy cereals according to protein, vitamin, sugar, and fat content, where healthy cereals should lie on the right part of the plot. The normalization procedure used in (b) shows that pointing healthy and unhealthy variables in opposite directions as in (a) only stretches and distorts the plot.
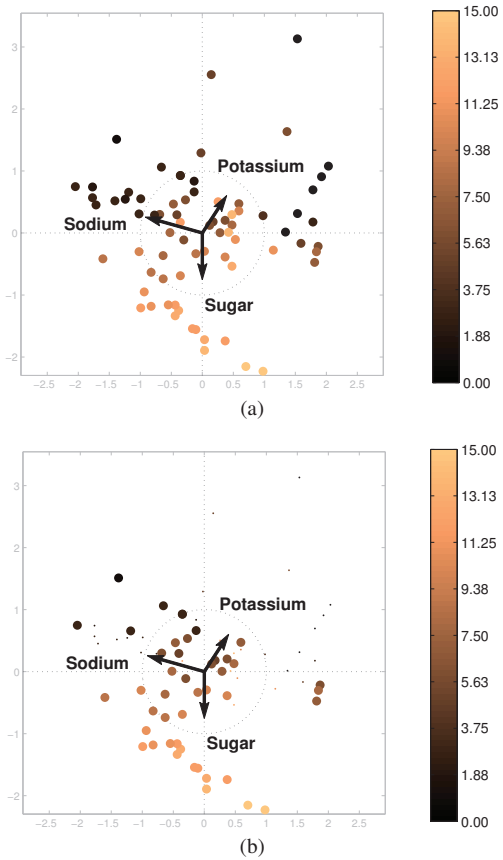

Fig. 7. Visualizing error information. A SC plot for 3 (standardized) breakfast cereal variables is shown in (a) as usual, for an orthonormalized configuration of axis vectors, where the colors represent sugar content. In (b), the samples for which the estimation errors are large are depicted as very small dots. In contrast, we can not only estimate values accurately for the remaining large dots, but the distances between them are almost identical to the ones in the data space.

ables in the opposite direction. Figure 6a shows a configuration of axis vectors used to characterize healthy vs. unhealthy breakfast cereals, similar to that in [24]. The result of using $\mathbf{V}_\perp$ with centered data, illustrated in Fig. 6b, shows that the original configuration unnecessarily stretched the plot along the horizontal axis, introducing distortions in the visualization.

In practice, the data samples will not lie on $\mathscr{R}(\mathbf{V})$ exactly, causing nonzero estimation errors that can be visualized in order to examine which data samples are being represented well for the particular set of axis vectors. When they are large they provide warnings that indicate the need to examine data samples carefully before making decisions. However, if $\mathbf{V}_\perp$ is used with centered data analysts can be confident not only about interpreting values of points that have been represented accurately, but also regarding distances and angles between them. This is illustrated in Fig. 7 through an example involving three breakfast cereal variables. In Fig. 7a the points are depicted as usual, where the color coding corresponds to sugar content. In Fig. 7b, we have decreased the size of the dots corresponding to samples for which the estimation errors are large. This operation simply shows samples that have not been represented well as very small dots, since they can be misleading. Notice that healthy cereals with high potassium values and low sugar and sodium have practically disappeared from the plot. This is useful for decision making, where the errors are warning us that the data set does not contain cereals with that combination of features. In addition, we can be very confident about the estimated values of the remaining larger dots, where the distances between them are almost identical to the ones in the data space.
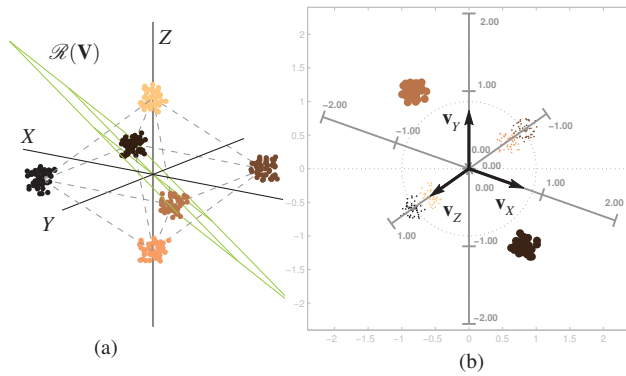
Fig. 8. Error information in cluster analysis. Image (a) shows an example with a 3-dimensional toy data set consisting in six small clusters centered at the vertices of an octahedron: $(-1,-1,0)$, $(-1,1,0)$, $(1,-1,0)$, $(1,1,0)$, $(0,0,1)$ and $(0,0,-1)$. The range of $\mathbf{V}$ only passes through two of the clusters, centered at $(1,-1,0)$ and $(-1,1,0)$, whose samples are represented well in the low-dimensional plot. The result of the linear mapping is shown in (b), where the points with larger estimation errors are depicted with smaller dots. Thus, analysts should be cautious when interpreting and comparing values or distances between those points.

Lastly, Fig. 8 shows an example with a 3-dimensional toy data set containing six small clusters. The particular choice of axis vectors results in the range of $\mathbf{V}$ passing through only two of the clusters, as shown in Fig. 8a. Therefore, the points in those clusters are represented well, and analysts can be confident about correctly interpreting and comparing their values. Furthermore, the relative distances between those points are also better represented. The result of the method's linear mapping is shown in Fig. 8b. In order to focus on the points that have been represented well, those with larger estimation errors have been displayed with smaller dots. Thus, it is important to note that the approach is effectively revealing the loss of information associated with the ability to recover original data values.

3 EVALUATION OF THE ESTIMATION ACCURACY

In this section we evaluate the accuracy of the estimates associated with the approaches on real data sets, and describe models of how users estimate data values in SC and RadViz. In particular, Tab. 1 summarizes the methods' approaches and estimation strategies used in the experiments.

The user study involved 10 computer science professors (all male, with average age 37) and 10 graduate students (with average age 28, where only 2 were female), at Universidad Rey Juan Carlos, in Madrid, Spain. The participants did not have a previous background in SC nor in RadViz.

3.1 Star Coordinates

Before performing estimation experiments we explained to them SC in detail. We made special emphasis on understanding the linear combination geometrically as a weighted sum of vectors, which could be interpreted as a path from the origin to the low-dimensional point $\mathbf{p}$, as illustrated in Fig. 1.

For this purpose, we developed a tool to teach the method prior to performing the estimation experiments. Given a configuration of axis vectors, and a low-dimensional point $\mathbf{p}$ corresponding to a particular data instance $\mathbf{x}$ from the breakfast cereal data set (with values in $[0,1]$), it allowed users to visualize and construct a path from the origin to $\mathbf{p}$ by updating the coefficients associated with the axis vectors. These values would represent possible estimates for the data sample $\mathbf{x}$ that had been mapped onto $\mathbf{p}$ by SC. At the end of the process the true values of $\mathbf{x}$ were revealed to the users. One of our goals consisted of showing that there is usually an infinite number of possibilities to choose the coefficients. Additionally, by using the cereal database, which is not sparse, we intended to show that in most cases the actual values of $\mathbf{x}$ were not 0 (i.e., most variables contributed a segment to the path). In total, each participant estimated values of nine data entries in this training stage: 3 (for sets of 3, 6, and 9 variables) × 3 instances.

Afterwards we measured the participants' ability to estimate data attributes in SC without the aid of calibrated axes. In a first experiment we showed users the embeddings of randomly chosen data samples from the breakfast cereal database and the wine data set in [6], together with configurations of axis vectors ($\mathbf{V}$) whose directions were random and lengths were drawn from a uniform $\mathcal{U}[0.5, 2.5]$. This particular choice implies that the longest vector will be at most 5 times larger than the smallest. On the one hand, we guarantee that the variable with the smallest vector will still have a relevant effect on the plot, since the contribution of a variable is proportional to the length of its axis vector. On the other hand, we chose the upper limit of 2.5 in order to expect a notable reduction of the estimation errors when carrying out the orthonormalization process. Note that $\mathbf{V}_\perp$ could be quite different than $\mathbf{V}$, since the length of its axis vectors can be at most 1, due to Prop. 3.

The experiments considered configurations of 3, 5, 7, and 9 variables, also chosen at random from the data sets. Finally, we asked the participants to estimate the value of only one of the attributes (also selected at random). In total, every participant estimated 40 values: 5 attribute values × 2 data sets × 4 (for sets of 3, 5, 7 and 9 variables). It is worth mentioning that we encouraged the participants to use a fast and intuitive estimation strategy that would not involve complex calculations, since otherwise users could simply look up the values of the data quicker on the graphical interface (e.g., in a table). This allowed them to report values typically in under 15 seconds.

Finally, we compared the results with those relative to automatic strategies based on projections onto calibrated axes (see Tab. 1). Figure 9 shows average estimation errors for the user (SC_{User}), and automatic estimates (see the supplemental material for separate results on each data set). The reported measure for a certain number of variables n is:

$$v(n) = \frac{1}{P \cdot D \cdot M} \sum_{p=1}^{P} \sum_{d=1}^{D} \sum_{i=1}^{M} |\hat{x}_{i,d,p} - x_{i,d,p}|,$$

where $P = 20$ is the number of participants, $D = 2$ is the number of data sets, $M = 5$ is the number of values to estimate, $x_{i,d,p}$ is the i-th value to be estimated for the d-th data set and p-th participant, and finally $\hat{x}_{i,d,p}$ is the corresponding estimate. It is apparent that, although the orthonormalization process has a positive effect, the estimates based on projections onto calibrated axes are only appropriate when the data is centered. In that case the approach clearly provides more accurate estimates than the users'. Without the aid of the calibrated axes the estimation task requires carrying out complex calculations that are generally inaccurate and too time consuming.

While in the first experiment users only had to estimate one value per data instance, we carried out another study where we asked the participants to estimate all of the attributes. Our goal was not only to evaluate the accuracy of users' estimates, but also to analyze their strategies.

In this new experiment we showed regular (with unit vectors) and random configurations of axis vectors (again with random directions
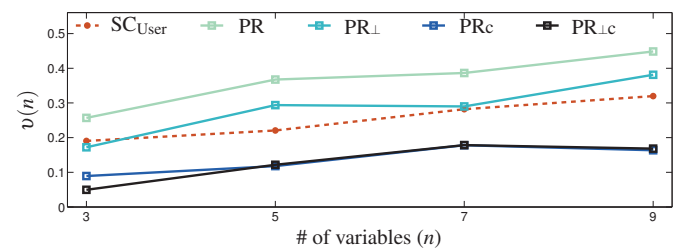

Fig. 9. Average estimation errors $v(n)$ obtained in experiments with users for SC, and the automatic estimates onto calibrated axes. Projecting points onto calibrated axes allows to obtain more accurate results, but the data must be previously centered.

Table 1. Summary of the Methods' Approaches and Estimation Strategies Used to Gather Results.

Result notation	Method	Data normalization	Axis vectors	Estimation strategy
SC_{User}	SC	[0,1]	$\mathbf{V}$	Reported by users
SC_M	SC	[0,1]	$\mathbf{V}$	Modeled by (6)
$RadViz_M$	RadViz	[0,1]	$\mathbf{V}$	Modeled by (7)
PR	SC	[0,1]	$\mathbf{V}$	Projections onto calibrated axes
$PR_\perp$	SC	[0,1]	$\mathbf{V}_\perp$	Projections onto calibrated axes
PR_c	SC	Range 1, centered	$\mathbf{V}$	Projections onto calibrated axes
$PR_{\perp c}$	SC	Range 1, centered	$\mathbf{V}_\perp$	Projections onto calibrated axes

and lengths drawn from a uniform $\mathscr{U}[0.5, 2.5]$), and the embedding of a randomly chosen data sample $\mathbf{x}$ from the breakfast cereal database, where the variables were normalized to lie in [0,1]. The experiments considered sets of 3, 6, and 9 variables, also chosen at random from the database. Finally, we asked the participants to estimate the attributes of $\mathbf{x}$. In total, every participant estimated the variable values of 12 data instances: 2 (for regular and random configurations) $\times$ 3 (for sets of 3, 6, and 9 variables) $\times$ 2 instances.

At the end of the experiment we asked the participants to explain their strategies. We identified one main approach, used by 13 participants (65%), where they searched for estimates (i.e., the coefficients of the linear combination) that minimize the length of the path from the origin to $\mathbf{p}$ by using the least number of variables as possible (typically, only two or three), while satisfying the linear combination equation (1), as well as the constraints associated with the data normalization. Thus, many of the reported estimates were 0, despite our efforts in the training stage to show that the solutions were usually not sparse. This simple approach arises not only due to the difficulty to estimate values, but also to the necessity to visualize them quickly. For this strategy users needed about 20 seconds to report the nonzero estimates. Note that the particular scale of the axis vectors does not affect the user estimates. In other words, we would have obtained identical results had we chosen a uniform $\mathscr{U}[0.2, 1]$ for the lengths of the axis vectors. Lastly, we did not identify other common strategies.

In order to extrapolate the results to more data we modeled the user strategy, denoted as SC_M, with the following convex optimization problem:

$$\begin{aligned} \text{minimize} \quad & \|\mathbf{Dx}\|_1 \\ \mathbf{x} \in \mathbb{R}^n \\ \text{subject to} \quad & \mathbf{V}^T\mathbf{x} = \mathbf{p}, \\ & \mathbf{0} \preceq \mathbf{x} \preceq \mathbf{1}, \end{aligned} \quad (6)$$

where $\mathbf{D}$ is diagonal matrix in which $d_{i,i} = \|\mathbf{v}_i\|$, and $\preceq$ denotes vector componentwise inequality. Note that $\|\mathbf{Dx}\|_1$ corresponds to the length of the path. Thus, the optimization problem models the fact that users generally search for solutions where the path associated with the linear combination is shortest. Moreover, the ℓ_1 norm is appropriate for obtaining solutions that use only a few variables (nonzero coefficients), since it usually produces sparse solutions, i.e., many values of $\hat{\mathbf{x}}$ will be 0, which is equivalent to not considering a variable in the linear combination. The supplemental material includes an alternative formulation (as a linear program) [1], and shows that the solutions are sparse.

Figure 10 analyzes the validity of model SC_M by examining the discrepancy between the model and user estimates (only for the 13 participants who used the associated strategy), and evaluating their average estimation errors. In Fig. 10a the estimates of the model ($\hat{\mathbf{x}}_{Model}$), are very close to the ones provided by the users ($\hat{\mathbf{x}}_{User}$). Thus, the average squared estimation errors for SC_{User} and SC_M are also very similar. In Fig. 10b the average estimation errors for random configurations are also very close, despite a higher discrepancy ($\|\hat{\mathbf{x}}_{User} - \hat{\mathbf{x}}_{Model}\|^2$) between the users' and the model's estimates.

3.2 RadViz

In addition, we studied possible ways users estimate values in RadViz. Since the method was new to the participants we explained to them its physical spring model and described its properties, as mentioned in Sec. 1.
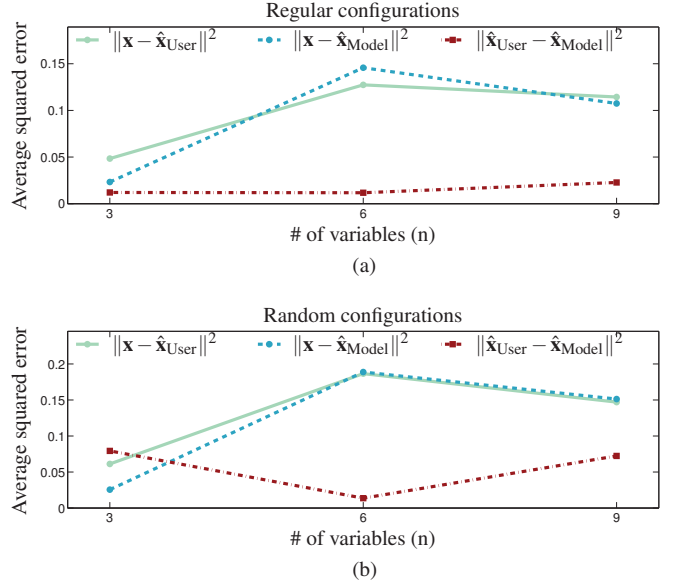

Fig. 10. Validity of SC_M for (a) regular, and (b) random configurations of axis vectors. In (a) the estimates of the model are very close to the ones provided by users ($\|\hat{\mathbf{x}}_{User} - \hat{\mathbf{x}}_{Model}\|^2$). Therefore, the average squared estimation errors for SC_{User} ($\|\mathbf{x} - \hat{\mathbf{x}}_{User}\|^2$) and SC_M ($\|\mathbf{x} - \hat{\mathbf{x}}_{Model}\|^2$) are also very similar. For random configurations there is more discrepancy between the model and the user estimates, but the average estimation errors are nevertheless still very similar.

Firstly, we carried out an experiment where, given a regular configuration of five variables (with anchor points distributed uniformly on the unit circle), and a point $\mathbf{p}$ inside the pentagon formed by the anchor points, users were asked to intuitively choose all of the values of some possible instance that would have been mapped onto $\mathbf{p}$. In particular, each participant chose the five attribute values associated with five points. Let IC denote these "intuitive choices".

Afterwards, we asked the participants to obtain a formula or algorithm for visualizing values quickly, which would be consistent with the RadViz properties. They were given 24 hours to come up with a strategy. The following day only two professors and one student reported a consistent strategy, which was identical. In particular, they formed estimates by tracing a line segment from the anchor $\mathbf{v}_i$ to $\mathbf{p}$, and finally reaches the convex hull of the anchors at a point $\mathbf{c}$. The i-th estimate $\hat{x}_i$ is then defined as:

$$\hat{x}_i = \|\mathbf{c} - \mathbf{p}\| / \|\mathbf{c} - \mathbf{v}_i\|. \quad (7)$$

This approach, denoted as $RadViz_M$, is illustrated in Fig. 11.

Other users reported strategies that were consistent with every property but the 4-th, when $\mathbf{p}$ is on an edge of the convex hull of the anchor points. For instance, 5 participants used $\hat{x}_i = \|\mathbf{c} - \mathbf{p}\|/d$, where $d = 2\|\mathbf{v}\|$ is the diameter of the circumference on which the anchor points lie. However, these users corrected their strategy to (7) when notified that their initial solution was inconsistent. Other solutions involved ratios of areas, or the ratio between the distance from anchor i
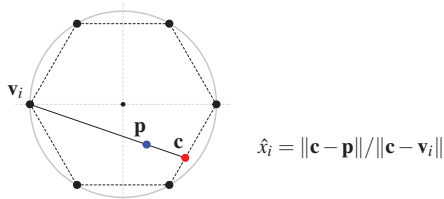
PSfrag replacements $\mathbf{v}_i$

$\hat{x}_i = \|\mathbf{c} - \mathbf{p}\|/\|\mathbf{c} - \mathbf{v}_i\|$

Fig. 11. Estimation strategy employed by users in RadViz for individual variables that is consistent with the method's properties.

to $\mathbf{p}$ and the sum of the distances from $\mathbf{p}$ to the rest of the anchors.

Finally, we compared the intuitive choices IC before the participants thought of a strategy, to those produced by the RadViz model for the same settings used in the initial experiment. Figure 12 shows a histogram of the differences between the estimates. Most of the discrepancies have an absolute value less than 0.1, which indicates that RadViz$_M$ approximates the users' intuitive estimates fairly well.

3.3 Comparison among models

We have empirically compared the estimation accuracy of the described approaches by carrying out a simulation involving the breakfast cereal data set, the Breast Cancer Wisconsin (Diagnostic), Parkinsons, and Wine data sets available in [6]. The experiment measured average estimation errors for configurations of axis vectors where the variables were selected at random. In particular the reported measure for a data set is:

$$\delta(n) = \frac{1}{n \cdot T} \sum_{t=1}^{T} \|\hat{\mathbf{x}}_t - \mathbf{x}_t\|,$$

where n is the number of variables, $T = 2000$ is the number of trials, $\mathbf{x}_t$ is the randomly chosen sample from the data set on the t-th trial, and $\hat{\mathbf{x}}_t$ is the corresponding vector of estimates for such sample.

Figure 13 shows results on random configurations of axis vectors. The vectors' angles were chosen at random, while their lengths followed a uniform $\mathscr{U}[0.5, 1]$ distribution. The figure shows that centering the data (PR$_c$) has a greater positive effect on estimation accuracy than the orthonormalization process alone (PR$_\perp$). However, both approaches combined lead to the best results (PR$_{\perp c}$). The results are poor for SC$_M$, PR, and even PR$_\perp$. In this regard, notice that the differences between $\mathbf{V}$ and $\mathbf{V}_\perp$ are relatively small, since the lengths of the axis vectors are at most 1 (see Prop. 3). Thus, the results for PR and PR$_\perp$, and for PR$_c$ and PR$_{\perp c}$, are similar. Finally, RadViz$_M$ can achieve, at most, the performance of PR$_c$. It is important to note that, since the axis vectors define the anchor points, in this setting these are not arranged in a regular pattern.

Figure 14 shows results on regular orthonormalized configurations of axis vectors. In these examples it is apparent that the centering is crucial for obtaining accurate estimates. In this case the anchors for RadViz are arranged in a regular pattern, which has a positive effect on the estimation accuracy of RadViz$_M$.

Figure 15 shows results on PCA configurations of axis vectors. In these cases the columns of $\mathbf{V}$, which are the eigenvectors of the data's
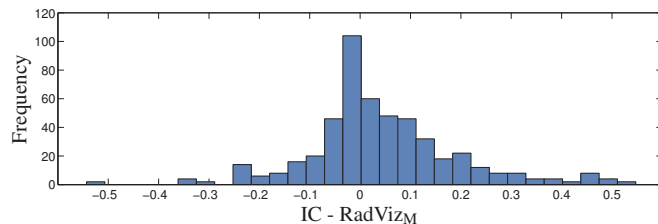

Fig. 12. Histogram of differences between the intuitive choices of values (IC) in the first RadViz user experiment, and the estimates generated by RadViz$_M$ for the same settings.

covariance matrix, form an orthonormal set, and therefore generate orthographic projections in SC. The examples also show the importance of centering regarding estimation accuracy. In all cases (Fig. 13, 14, and 15), the most accurate estimates are obtained by PR$_{\perp c}$ (orthonormalized vectors with centered data), as predicted by the theoretical analysis.

4 DISCUSSION

In this paper we have addressed one of the main drawbacks of SC, namely, the difficulty to estimate data attributes, providing a new approach to mitigate the issue where estimates are computed through projections onto calibrated axes, similarly to statistical biplots. However, the data and axis vectors must satisfy certain conditions in order to obtain accurate estimates. In particular, the columns of $\mathbf{V}$ must form an orthonormal set of vectors, and the data must be centered.

Specifically, the estimates after centering the data can be considerably more accurate than those provided by orthonormalizing the columns of $\mathbf{V}$, showing that normalizing the data to lie in $[0, 1]$ is a poor choice regarding estimation accuracy when using projections onto calibrated axes. However, both approaches can be combined to enhance the estimates even further, where the orthonormalization provides several benefits, since it allows to represent the data more faithfully. Additionally, we have also shown that the estimates in our experiments with $\mathbf{V}_\perp$ and centered data are more accurate than those obtained through the SC and RadViz estimation models.

Another important contribution is a detailed geometrical description of the decomposition of the dimensionality reduction mapping, which explains the optimality of the approaches, why distances in $\mathscr{R}(\mathbf{V})$ are preserved on the display, and how to understand plotted information and estimation errors.

The axis calibration also allows to visualize the errors and interpret them geometrically. This is useful for data analysis since it provides a measure of goodness of a plot (see [2]). In particular, the errors can be used to discard or highlight data samples that have not been represented well and can be misleading, which can enhance decision support tasks. In addition, they can improve our interpretation of the distances between the embedded samples.

Naturally, since there is a loss of information in any dimensionality reduction process, our goal is not to recover exact values, but to obtain better estimates for exploratory purposes. Note that the ability to represent an entire data set accurately depends on how well $\mathscr{R}(\mathbf{V})$ fits the data. Unless the points lie close to an m-dimensional linear manifold, and the axis vectors are chosen so that $\mathscr{R}(\mathbf{V})$ is aligned with it, the estimates for some points will inevitably be inaccurate. In this regard, the configuration of axis vectors for which $\mathscr{R}(\mathbf{V})$ fits the data optimally (i.e., minimizes the sum of squared estimation errors over every sample) leads to the PCA plot. However, while in PCA the axis vectors are fixed, SC allows to manipulate them interactively, which is useful in other tasks such as searching for cluster structure, outlier detection, or searching for data with particular characteristics.

Finally, these improvements on attribute estimation are focused on addressing one of the major goals of data visualization, which is to provide meaningful representations of data that will allow analysts to extract knowledge visually, without the need to inspect concrete values in tables.

REFERENCES

[1] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.
[2] C. Chen. Top 10 unsolved information visualization problems. *IEEE Comput. Graph. Appl.*, 25:12–16, July 2005.
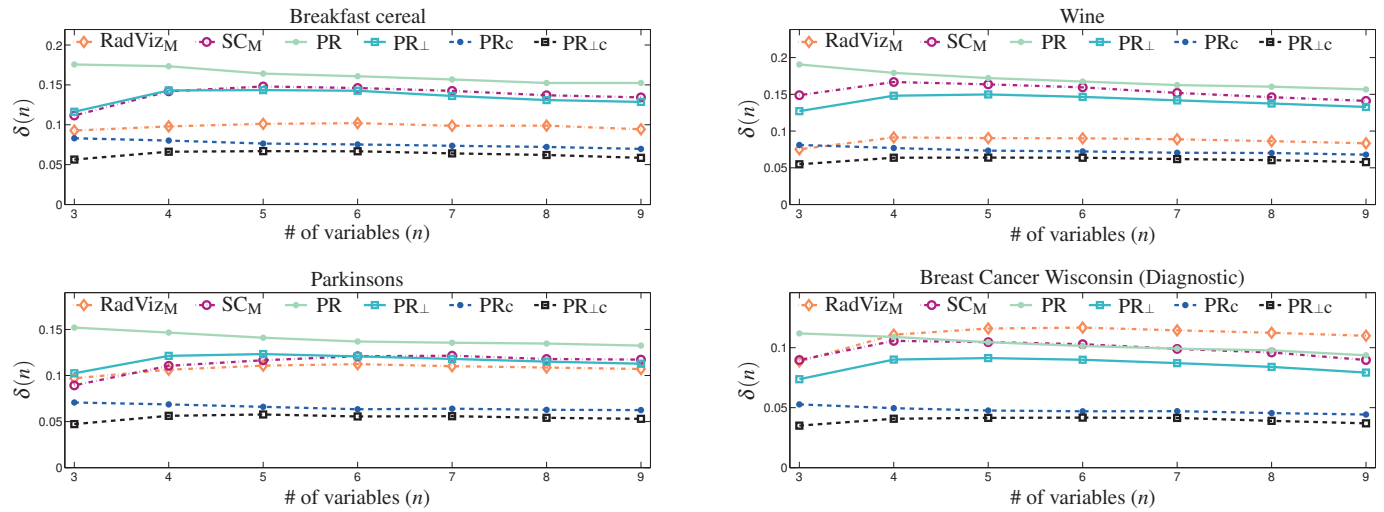
replacements

Fig. 13. Average estimation error $\delta(n)$ obtained in a simulation for random configurations of axis vectors.
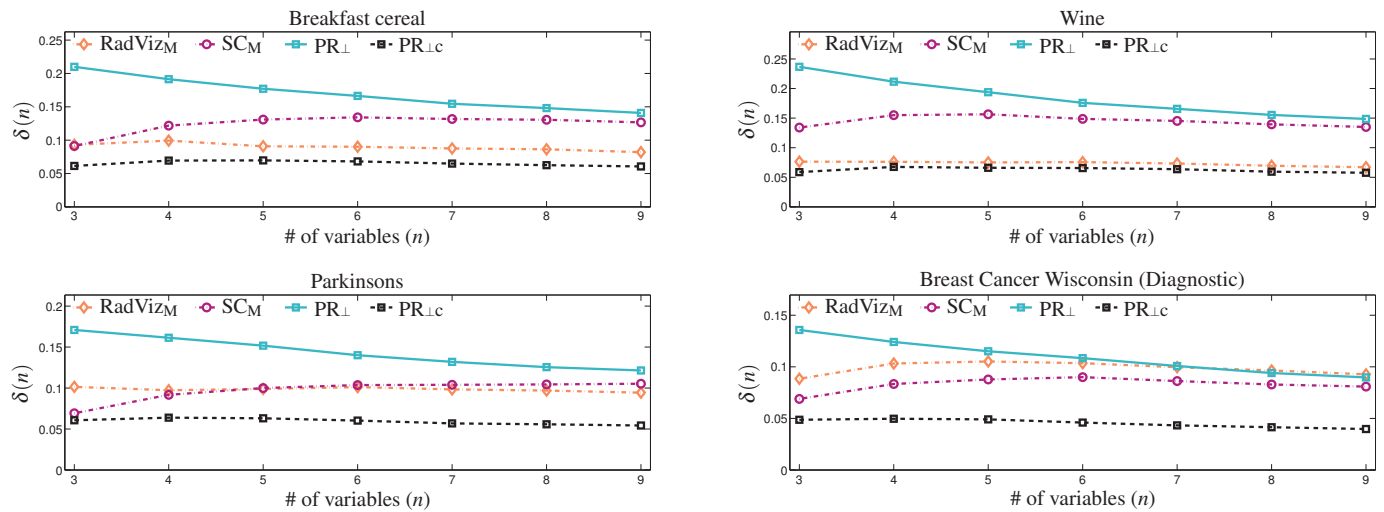
Fig. 14. Average estimation error $\delta(n)$ obtained in a simulation for regular configurations $\mathbf{V}_\perp$ of axis vectors, which lead to orthographic projections in SC. Note that in this case PR = PR$_\perp$, and PRc = PR$_\perp$c.
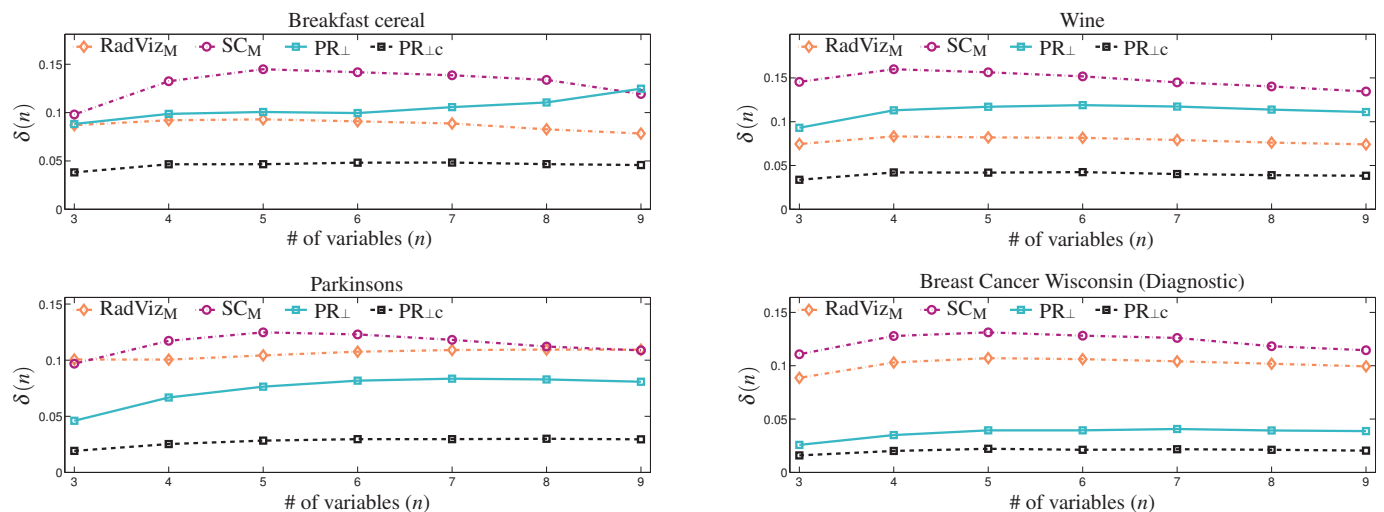
Fig. 15. Average estimation error $\delta(n)$ obtained in a simulation for PCA configurations $\mathbf{V}_\perp$ of axis vectors, which form orthographic projections in SC. Note that in this case PR = PR$_\perp$, and PRc = PR$_\perp$c.

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

[4] S. Diehl, F. Beck, and M. Burch. Uncovering strengths and weaknesses of radial visualizations—an empirical approach. *IEEE Transactions on Visualization and Computer Graphics*, 16:935–942, November 2010.

[5] G. M. Draper, Y. Livnat, and R. F. Riesenfeld. A survey of radial methods for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 15:759–776, September 2009.

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

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

[8] J. Gower, S. Gardner-Lubbe, and N. le Roux. *Understanding Biplots*. John Wiley & Sons, 2011.

[9] M. Greenacre. *Biplots in Practice*. BBVA Foundation, 2010.

[10] 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.

[11] A. Inselberg. *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2009.

[12] A. Inselberg and B. Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *Proceedings of the 1st conference on Visualization*, VIS'90, pages 361–378, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.

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

[14] 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.

[15] 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.

[16] D. A. Keim. Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8:1–8, January 2002.

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

[18] T. V. Long and L. Linsen. Visualizing high density clusters in multidimensional data using optimized star coordinates. *Computational Statistics*, 26(4):655–678, 2011.

[19] R. Rao and S. K. Card. The table lens: merging graphical and symbolic representations in an interactive focus+context visualization for tabular information. In *Conference companion on Human factors in computing systems*, CHI'94, pages 318–322, New York, NY, USA, 1994. ACM.

[20] 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.

[21] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*, pages 336–343, Washington, DC, USA, 1996. IEEE Computer Society.

[22] H. Siirtola and K.-J. Räihä. Interacting with parallel coordinates. *Interacting with Computers*, 18(6):1278 – 1309, 2006.

[23] 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.

[24] 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.

[25] K.-B. Zhang, M. Orgun, and K. Zhang. HOV[3]: An approach to visual cluster analysis. In *Advanced Data Mining and Applications*, volume 4093 of *Lecture Notes in Computer Science*, pages 316–327. Springer Berlin / Heidelberg, 2006.