

Appendix A: Supplemental material

Optimal calibration scalings and shifts

Let f denote the objective function in (9). Firstly, setting the partial derivative of f with respect to β_i equal to 0 we have:

$$\frac{\partial f}{\partial \beta_i} = 2 \sum_{j=1}^N (\alpha_j \mathbf{p}_j^\top \mathbf{v}_i + \beta_i - x_{j,i}) = 0.$$

Solving for β_i yields (11):

$$\beta_i = \bar{x}_i - \frac{\alpha_i}{N} \sum_{j=1}^N \mathbf{p}_j^\top \mathbf{v}_i.$$

Secondly, setting the partial derivative of f with respect to α_i equal to 0 we have:

$$\frac{\partial f}{\partial \alpha_i} = 2 \sum_{j=1}^N (\alpha_j \mathbf{p}_j^\top \mathbf{v}_i + \beta_i - x_{j,i}) \mathbf{p}_j^\top \mathbf{v}_i = 0.$$

Simplifying and incorporating the expression for β_i we have:

$$\alpha_i \sum_{j=1}^N (\mathbf{p}_j^\top \mathbf{v}_i)^2 + \left(\bar{x}_i - \frac{\alpha_i}{N} \sum_{j=1}^N \mathbf{p}_j^\top \mathbf{v}_i \right) \left(\sum_{j=1}^N \mathbf{p}_j^\top \mathbf{v}_i \right) - \sum_{j=1}^N x_{j,i} (\mathbf{p}_j^\top \mathbf{v}_i) = 0.$$

Finally, solving for α_i yields (10):

$$\alpha_i = \frac{\sum_{j=1}^N x_{j,i} (\mathbf{p}_j^\top \mathbf{v}_i) - \bar{x}_i \sum_{j=1}^N \mathbf{p}_j^\top \mathbf{v}_i}{\sum_{j=1}^N (\mathbf{p}_j^\top \mathbf{v}_i)^2 - \frac{1}{N} \left(\sum_{j=1}^N \mathbf{p}_j^\top \mathbf{v}_i \right)^2}.$$

Identical approximation accuracy for OSC and ARA when applying CAL

Proposition 1 Consider applying CAL on an OSC plot with matrix \mathbf{V}_\perp , and an ARA plot with matrix \mathbf{V} . If \mathbf{V}_\perp and \mathbf{V} span the same subspace (i.e., if $\mathcal{R}(\mathbf{V}) = \mathcal{R}(\mathbf{V}_\perp)$), the estimates $\hat{x}_{j,i} = \alpha_i (\mathbf{p}_j^\top \mathbf{v}_i) + \beta_i$ are identical in both plots, where \mathbf{v}_i denotes the i -th axis vector in either method.

Proof The proposition holds since the dot products $\mathbf{p}_j^\top \mathbf{v}_i$ are the same in both methods, which also implies that CAL will find identical values of α_i and β_i for a given data set \mathbf{X} . The values $\mathbf{p}_j^\top \mathbf{v}_i$ are the entries of the vector of dot products $\mathbf{V}\mathbf{p}$, which is the orthogonal projection of the data sample \mathbf{x}_j onto $\mathcal{R}(\mathbf{V})$, and therefore identical in both methods.

For example, in ARA we have:

$$\mathbf{V}\mathbf{p} = \mathbf{V}(\mathbf{V}^\top \mathbf{V})^{-1} \mathbf{V}^\top \mathbf{x},$$

while in OSC:

$$\mathbf{V}_\perp \mathbf{p} = \mathbf{V}_\perp \mathbf{V}_\perp^\top \mathbf{x} = \mathbf{V}_\perp (\mathbf{V}_\perp^\top \mathbf{V}_\perp)^{-1} \mathbf{V}_\perp^\top \mathbf{x}.$$

Recall that $\mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{x}$ is the orthogonal projection of \mathbf{x} onto $\mathcal{R}(\mathbf{A})$. Since we have assumed that $\mathcal{R}(\mathbf{V}) = \mathcal{R}(\mathbf{V}_\perp)$ it follows that $\mathbf{V}\mathbf{p} = \mathbf{V}_\perp \mathbf{p}$. \square

Solutions for optimal axes

In this section we show that the solutions to (13) are given by (14) and (15).

Firstly, the objective function in (13) can be rewritten as:

$$\begin{aligned} f_i(\mathbf{v}_i, \gamma_i) &= \sum_{j=1}^N (\mathbf{p}_j^\top \mathbf{v}_i + \gamma_i - x_{j,i})^2 \\ &= \|\mathbf{P}\mathbf{v}_i + \gamma_i \mathbf{1} - \mathbf{x}_i\|^2 \\ &= (\mathbf{P}\mathbf{v}_i + \gamma_i \mathbf{1} - \mathbf{x}_i)^\top (\mathbf{P}\mathbf{v}_i + \gamma_i \mathbf{1} - \mathbf{x}_i) \\ &= \mathbf{v}_i^\top \mathbf{P}^\top \mathbf{P} \mathbf{v}_i - 2\mathbf{v}_i^\top \mathbf{P}^\top \mathbf{x}_i + 2\gamma_i \mathbf{v}_i^\top \mathbf{P}^\top \mathbf{1} \\ &\quad + N\gamma_i^2 - 2\gamma_i \mathbf{x}_i^\top \mathbf{1} + \mathbf{x}_i^\top \mathbf{x}_i, \end{aligned}$$

where \mathbf{P} is the $N \times 2$ matrix of plotted points (not necessarily centered), $\mathbf{1}$ is a vector of N ones, and \mathbf{x}_i is the N -dimensional vector of attribute values for the i -th data variable.

The partial derivatives with respect to γ_i and \mathbf{v}_i are:

$$\frac{\partial f_i}{\partial \gamma_i} = -2\mathbf{v}_i^\top \mathbf{P}^\top \mathbf{1} + 2N\gamma_i + 2\mathbf{x}_i^\top \mathbf{1}, \quad (23)$$

and

$$\frac{\partial f_i}{\partial \mathbf{v}_i} = 2\mathbf{P}^\top \mathbf{P} \mathbf{v}_i - 2\mathbf{P}^\top \mathbf{x}_i - 2\gamma_i \mathbf{P}^\top \mathbf{1}. \quad (24)$$

Setting (23) to 0 yields:

$$\begin{aligned} \gamma_i &= \frac{1}{N} (\mathbf{x}_i^\top \mathbf{1} - \mathbf{v}_i^\top \mathbf{P}^\top \mathbf{1}) = \frac{1}{N} \mathbf{1}^\top (\mathbf{x}_i - \mathbf{P}\mathbf{v}_i) \\ &= \frac{1}{N} \sum_{j=1}^N (x_{j,i} - \mathbf{p}_j^\top \mathbf{v}_i) = \bar{x}_i - \frac{1}{N} \sum_{j=1}^N \mathbf{p}_j^\top \mathbf{v}_i. \end{aligned} \quad (25)$$

Substituting the expression for γ_i in (24) and setting the partial derivative to 0 yields:

$$\mathbf{P}^\top \mathbf{P} \mathbf{v}_i - \mathbf{P}^\top \mathbf{x}_i - \frac{1}{N} \mathbf{1}^\top (\mathbf{P}\mathbf{v}_i - \mathbf{x}_i) \mathbf{P}^\top \mathbf{1} = \mathbf{0}.$$

Since $\mathbf{1}^\top \mathbf{P}\mathbf{v}_i$ and $\mathbf{1}^\top \mathbf{x}_i$ are scalars we can write the equation as:

$$\mathbf{P}^\top \mathbf{P} \mathbf{v}_i - \mathbf{P}^\top \mathbf{x}_i - \frac{1}{N} \mathbf{P}^\top \mathbf{1} \mathbf{1}^\top \mathbf{P} \mathbf{v}_i + \frac{1}{N} \mathbf{P}^\top \mathbf{1} \mathbf{1}^\top \mathbf{x}_i = \mathbf{0},$$

$$\mathbf{P}^\top \left(\mathbf{I} - \frac{1}{N} \mathbf{1} \mathbf{1}^\top \right) \mathbf{P} \mathbf{v}_i = \mathbf{P}^\top \left(\mathbf{I} - \frac{1}{N} \mathbf{1} \mathbf{1}^\top \right) \mathbf{x}_i,$$

where \mathbf{I} is the $N \times N$ identity matrix. Additionally, $\mathbf{I} - (1/N)\mathbf{1}\mathbf{1}^\top$ is the well-known ‘‘centering’’ matrix, which is symmetric and idempotent. Thus, we can rewrite the previous equation as:

$$\mathbf{P}_c^\top \mathbf{P}_c \mathbf{v}_i = \mathbf{P}_c^\top \mathbf{x}_i,$$

where $\mathbf{P}_c = (\mathbf{I} - (1/N)\mathbf{1}\mathbf{1}^\top) \mathbf{P}$ is the centered version of \mathbf{P} (i.e., its column sums are 0). Also, note that the data samples can also be centered, in which case the embedded points of any linear transformation will also be centered. Finally, assuming \mathbf{P}_c has rank 2 we have:

$$\mathbf{v}_i = (\mathbf{P}_c^\top \mathbf{P}_c)^{-1} \mathbf{P}_c^\top \mathbf{x}_i = \mathbf{P}_c^+ \mathbf{x}_i.$$

Identical approximation accuracy for SC, OSC, and ARA when applying OPT

Proposition 2 Let \mathbf{X} represent an $N \times n$ data matrix, \mathbf{V} an $n \times 2$ matrix of full column rank, and \mathbf{M} a 2×2 invertible matrix. In addition, let \mathbf{P} denote the $N \times 2$ matrix that is the result of mapping the N data samples of \mathbf{X} linearly onto a plane through the matrix \mathbf{VM} . In other words:

$$\mathbf{P} = \mathbf{XVM}. \quad (26)$$

Furthermore, let $\mathbf{V}_* = \mathbf{P}^\dagger \mathbf{X}$ denote a matrix of enhanced axis vectors as defined in (16). Finally, let \mathbf{PV}_*^\top represent approximations of data samples in \mathbf{X} by projecting the N embedded points in \mathbf{P} orthogonally onto the enhanced axes, as defined in biplots or ARA plots (i.e., the approximations are the dot products between the embedded points and the enhanced axis vectors). In that case, the approximations \mathbf{PV}_*^\top do not depend on \mathbf{M} .

Proof Firstly, $\mathbf{V}_* = \mathbf{P}^\dagger \mathbf{X} = [\mathbf{XVM}]^\dagger \mathbf{X}$. Thus, we can express the approximations as:

$$\mathbf{PV}_*^\top = \mathbf{XVM}(\mathbf{M}^\top \mathbf{V}^\top \mathbf{X}^\top \mathbf{XVM})^{-1} \mathbf{M}^\top \mathbf{V}^\top \mathbf{X}^\top \mathbf{X}.$$

Since $(\mathbf{AB})^{-1} = \mathbf{B}^{-1} \mathbf{A}^{-1}$ for $m \times m$ invertible square matrices \mathbf{A} and \mathbf{B} , we can rewrite the approximations as:

$$\begin{aligned} \mathbf{PV}_*^\top &= \mathbf{XVMM}^{-1}(\mathbf{V}^\top \mathbf{X}^\top \mathbf{XV})^{-1}(\mathbf{M}^\top)^{-1} \mathbf{M}^\top \mathbf{V}^\top \mathbf{X}^\top \mathbf{X} \\ &= \mathbf{XV}(\mathbf{V}^\top \mathbf{X}^\top \mathbf{XV})^{-1} \mathbf{V}^\top \mathbf{X}^\top = (\mathbf{XV})(\mathbf{XV})^\dagger \mathbf{X}. \end{aligned}$$

Thus, they do not depend on matrix \mathbf{M} . \square

Corollary 1 Let \mathbf{X} represent an $N \times n$ data matrix, \mathbf{V} an $n \times 2$ matrix of full column rank, and \mathbf{V}_\perp and orthogonal matrix with the same range as \mathbf{V} . In addition, consider mapping the data samples in \mathbf{X} onto a plane with SC, ARA and OSC, through (2), (7), and (3), respectively. The approximations of \mathbf{X} resulting from projecting the embedded points orthogonally onto enhanced labeled axes, as they are defined in biplots or ARA plots, and which are obtained through (16), are identical for the three methods.

Proof The mappings for SC, ARA and OSC all have the form in (26). In particular, for SC $\mathbf{M} = \mathbf{I}$, for ARA $\mathbf{M} = (\mathbf{V}^\top \mathbf{V})^{-1}$, and for OSC $\mathbf{M} = \mathbf{B}$, where $\mathbf{V}_\perp = \mathbf{VB}$. Therefore, due to Prop. 2, the approximations when using the enhanced axes are identical for the three methods. \square

Solution for θ^*

We now show that the solution to (17) is given by (18). Firstly, recall that $\mathbf{P} = \mathbf{XV}$ is the set of embedded points prior to performing the scaling by θ , and $\mathbf{P}_c = \mathbf{CXV}$ contains the corresponding centered points, where $\mathbf{C} = \mathbf{I} - (1/N)\mathbf{1}\mathbf{1}^\top$ is the symmetric and idempotent centering matrix. Similarly, we denote the set of embedded points after performing the scaling as $\mathbf{P}^\theta = \theta\mathbf{XV}$, while $\mathbf{P}_c^\theta = \theta\mathbf{CXV}$ is its centered version.

The objective function of the optimization problem can be writ-

ten as:

$$\begin{aligned} f(\theta) &= \|\theta\mathbf{V} - \mathbf{V}_*^\theta\|_{\mathbb{F}}^2 = \|\theta\mathbf{V}^\top - (\mathbf{P}_c^\theta)^\dagger \mathbf{X}\|_{\mathbb{F}}^2 \\ &= \|\theta\mathbf{V}^\top - (\theta\mathbf{CXV})^\dagger \mathbf{X}\|_{\mathbb{F}}^2 \\ &= \|\theta\mathbf{V}^\top - (\theta^2 \mathbf{V}^\top \mathbf{X}^\top \mathbf{C}^2 \mathbf{XV})^{-1} \theta \mathbf{V}^\top \mathbf{X}^\top \mathbf{CX}\|_{\mathbb{F}}^2 \\ &= \|\theta\mathbf{V}^\top - \frac{1}{\theta} \mathbf{P}_c^\dagger \mathbf{X}\|_{\mathbb{F}}^2 = \|\theta\mathbf{V}^\top - \frac{1}{\theta} \mathbf{V}_*^\top\|_{\mathbb{F}}^2 \\ &= \text{tr} \left[\left(\theta\mathbf{V}^\top - \frac{1}{\theta} \mathbf{V}_*^\top \right)^\top \left(\theta\mathbf{V}^\top - \frac{1}{\theta} \mathbf{V}_*^\top \right) \right] \\ &= \theta^2 \text{tr}(\mathbf{V}^\top \mathbf{V}) - 2\text{tr}(\mathbf{V}_*^\top \mathbf{V}) + \frac{1}{\theta^2} \text{tr}(\mathbf{V}_*^\top \mathbf{V}_*) \\ &= \theta^2 \|\mathbf{V}\|_{\mathbb{F}}^2 - 2\text{tr}(\mathbf{V}_*^\top \mathbf{V}) + \frac{1}{\theta^2} \|\mathbf{V}_*\|_{\mathbb{F}}^2, \end{aligned}$$

where tr denotes the trace of a matrix. Setting its derivative equal to zero yields:

$$f'(\theta) = 2\theta \|\mathbf{V}\|_{\mathbb{F}}^2 - \frac{2}{\theta^3} \|\mathbf{V}_*\|_{\mathbb{F}}^2 = 0.$$

Finally, solving for θ we have:

$$\theta^* = \sqrt[4]{\frac{\|\mathbf{V}_*\|_{\mathbb{F}}^2}{\|\mathbf{V}\|_{\mathbb{F}}^2}} = \sqrt{\frac{\|\mathbf{V}_*\|_{\mathbb{F}}}{\|\mathbf{V}\|_{\mathbb{F}}}}.$$

Relationship between accuracy and axis vector length in ARA

There is a direct relationship between approximation accuracy and axis vector length in ARA plots. Since integers on the i -th line axis are located at multiples of $1/\|\mathbf{v}_i\|$, they appear closer to each other for larger axis vectors. This implies that a variation in \mathbf{p} in the direction of a large axis vector will cause a larger approximation error for the variable. Thus, the method will primarily focus on minimizing the approximation errors for variables with larger axis vectors. In Fig. 9 we illustrate this effect by comparing an initial ARA plot to another in which we have enlarged one axis vector. The example is based on the standardized Breakfast cereal data set used in [YMSJ05], but have labeled the axes with original data values. In this case, when an axis vector associated with Calories is stretched the plotted points appear more compacted in the direction of the axis (in SC the effect would be the opposite). Furthermore, the corresponding approximations are more accurate. Thus, the plotted points appear better ordered in the direction of the axis, as can be seen through the color coding of the dots, which represents caloric content.

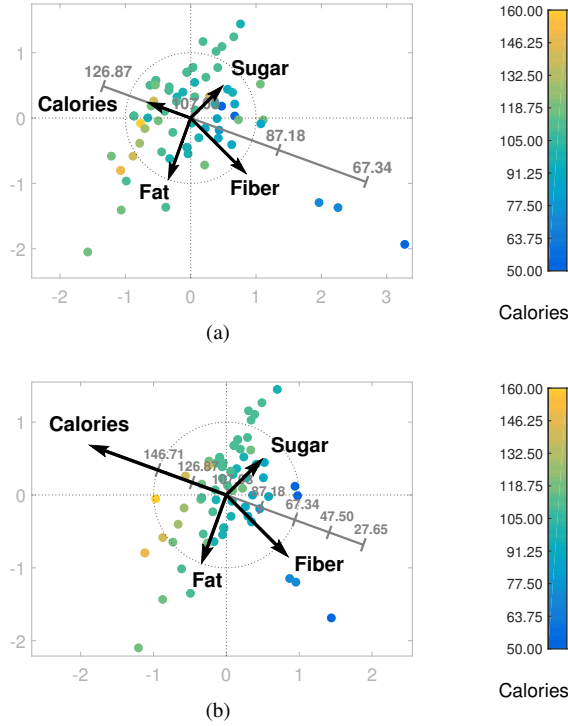


Figure 9: Direct relationship between accuracy and axis vector length in ARA plots. In this example the only difference between the ARA plots in (a) and (b) is that the axis vector associated with the variable *Calories* is longer in the latter. This compresses the plotted points along the direction of the axis in (b), and improves the approximation accuracy for *Calories*. In particular, observe (through the color coding) that the points appear better ordered with respect to caloric content along the direction of the axis.

Decomposition of the estimation errors ε_i

The objective function in (13), denoted here as ε_i can be rewritten as follows:

$$\begin{aligned}
 \varepsilon_i &= \sum_{j=1}^N (\mathbf{p}_j^T \mathbf{v}_i^* + \gamma_i^* - x_{j,i})^2 = \|\mathbf{P} \mathbf{v}_i^* + \mathbf{1} \gamma_i^* - \mathbf{x}_i\|^2 \\
 &= \left\| \mathbf{P} \mathbf{P}_c^\dagger \mathbf{x}_i + \mathbf{1} \left(\frac{1}{N} \mathbf{1}^T \mathbf{x}_i - \frac{1}{N} \mathbf{1}^T \mathbf{P} \mathbf{P}_c^\dagger \mathbf{x}_i \right) - \mathbf{x}_i \right\|^2 \\
 &= \left\| \left(\mathbf{I} - \frac{1}{N} \mathbf{1} \mathbf{1}^T \right) \mathbf{P} \mathbf{P}_c^\dagger \mathbf{x}_i - \left(\mathbf{I} - \frac{1}{N} \mathbf{1} \mathbf{1}^T \right) \mathbf{x}_i \right\|^2 \\
 &= \|\mathbf{P}_c \mathbf{P}_c^\dagger - \mathbf{C}\| \mathbf{x}_i\|^2 \\
 &= \mathbf{x}_i^T (\mathbf{P}_c \mathbf{P}_c^\dagger - \mathbf{C})^T (\mathbf{P}_c \mathbf{P}_c^\dagger - \mathbf{C}) \mathbf{x}_i \\
 &= \mathbf{x}_i^T (\mathbf{C} - \mathbf{P}_c \mathbf{P}_c^\dagger) \mathbf{x}_i = \mathbf{x}_i^T \mathbf{C} \mathbf{x}_i - \mathbf{x}_i^T \mathbf{P}_c \mathbf{P}_c^\dagger \mathbf{x}_i \\
 &= N \sigma_i^2 - \mathbf{x}_i^T \mathbf{C} \mathbf{P}_c \mathbf{P}_c^\dagger \mathbf{C} \mathbf{x}_i \\
 &= N \sigma_i^2 - \mathbf{x}_{c,i}^T \mathbf{P}_c \mathbf{P}_c^\dagger \mathbf{x}_{c,i}.
 \end{aligned}$$

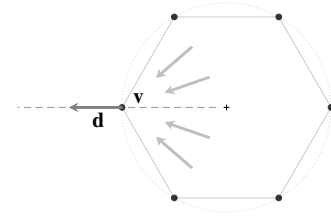


Figure 10: In RadViz increasing the value of a data variable causes the plotted points to move towards the anchor point \mathbf{v} associated with the variable. On average we assume that we should expect to find greater values for the variable in the direction (\mathbf{d}) from the origin towards \mathbf{v} .

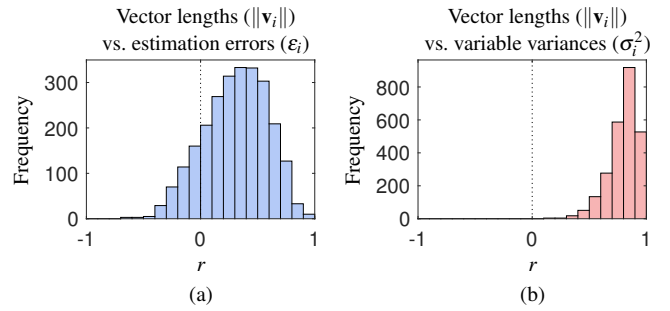


Figure 11: Histograms of Pearson correlations (r) between the length of an optimal axis vector and the estimation error, and the variance, of the associated variable, for the data used in the RadViz example in Sec. 3.2.4. The variables were first normalized to lie in $[0, 1]$, and afterwards for each sample we divided each attribute by the sum of all of the attributes. Since each variable has a different variance, $\|\mathbf{v}_i\|$ is usually positively correlated with σ_i^2 , but not necessarily with ε_i .

Effect of increasing an attribute in RadViz

Figure 10 shows the effect of increasing an attribute value in RadViz. The plotted point moves towards the anchor associated with the corresponding variable. We assume that on average the data values for a variable should increase in the direction (\mathbf{d}) from the origin to the anchor point \mathbf{v} .

Correlations related to $\|\mathbf{v}_i\|$ for the RadViz example

In the RadViz example (see Sec. 3.2.4) the length of the optimal vectors predominantly reflects the variance of the variables. In Fig. 11 we show distributions of correlations between $\|\mathbf{v}_i\|$, and ε_i and σ_i^2 , for the $7!/2$ different orderings of the eight variables (discarding rotations and reflections). The vector lengths usually have a strong positive correlation with the variable variances. In this example the variance for Palmitic is 0.0009, which is at least twice as small as the rest of the variances, which explains the short length of the Palmitic axis vector in Fig. 6.

Discarding offset shifts in the objective functions

The objective functions of the optimization problems considered in this paper have the following form:

$$f = \|\mathbf{P}\mathbf{V}^T - \mathbf{X} + \mathbf{1}\delta^T\|_{\mathbb{F}}^2, \quad (27)$$

where the variables related to the axis vectors appear in \mathbf{V} and \mathbf{P} . Also, assume \mathbf{V} , \mathbf{P} , and δ are or contain optimal solutions. In that case δ is (see the derivation of (25)):

$$\delta = \bar{\mathbf{x}} - \mathbf{V}\bar{\mathbf{p}} = -(\mathbf{V}\mathbf{W} - \mathbf{I})\bar{\mathbf{x}},$$

where $\bar{\mathbf{x}}$ and $\bar{\mathbf{p}}$ are the mean of the data and plotted points, respectively. Additionally, $\mathbf{W} = \mathbf{V}^T$ for SC, and $\mathbf{W} = \mathbf{V}^\dagger$ for ARA. Substituting in (27) we can rewrite the objective function as:

$$\begin{aligned} f &= \|\mathbf{V}\mathbf{P}^T - \mathbf{X}^T + \delta\mathbf{1}^T\|_{\mathbb{F}}^2 \\ &= \|\mathbf{V}\mathbf{W}\mathbf{X}^T - \mathbf{X}^T - (\mathbf{V}\mathbf{W} - \mathbf{I})\bar{\mathbf{x}}\mathbf{1}^T\|_{\mathbb{F}}^2 \\ &= \|(\mathbf{V}\mathbf{W} - \mathbf{I})\mathbf{X}^T - (\mathbf{V}\mathbf{W} - \mathbf{I})\bar{\mathbf{x}}\mathbf{1}^T\|_{\mathbb{F}}^2 \\ &= \|(\mathbf{V}\mathbf{W} - \mathbf{I})(\mathbf{X}^T - \bar{\mathbf{x}}\mathbf{1}^T)\|_{\mathbb{F}}^2. \end{aligned}$$

If we apply a translation \mathbf{s} to the data, the new data matrix would become $\mathbf{X} + \mathbf{1}\mathbf{s}^T$, while the new mean would be $\bar{\mathbf{x}} - \mathbf{s}$. In that case, f would not change:

$$\begin{aligned} f &= \|(\mathbf{V}\mathbf{W} - \mathbf{I})(\mathbf{X}^T + \mathbf{s}\mathbf{1}^T - (\bar{\mathbf{x}} - \mathbf{s})\mathbf{1}^T)\|_{\mathbb{F}}^2 \\ &= \|(\mathbf{V}\mathbf{W} - \mathbf{I})(\mathbf{X}^T - \bar{\mathbf{x}}\mathbf{1}^T)\|_{\mathbb{F}}^2. \end{aligned}$$

Thus, we obtain the same value for the objective function using centered data:

$$f = \|\mathbf{W}\mathbf{V}^T - \mathbf{X}_c + \mathbf{1}\delta^T\|_{\mathbb{F}}^2.$$

However, for centered data $\delta^T = \mathbf{0}$. Thus, the optimum value of the objective function is:

$$f = \|\mathbf{W}\mathbf{V}^T - \mathbf{X}_c\|_{\mathbb{F}}^2,$$

which implies that we obtain the same optimum axis vectors (as in (27)) solving the optimization problems on centered data but discarding the term involving δ .

Optimal scaling of a single axis vector

Table. 2 shows the derivation of the solution to (20) for SC.

Gradient of the objective function in (22) for SC

Table. 3 shows the derivation of the gradient of the objective function (f) in (22) for SC.

The objective function in (20) for SC can be rewritten as follows:

$$\begin{aligned}
P(\lambda) &= \sum_{j=1}^N \left\| \begin{bmatrix} \tilde{\mathbf{V}} \\ \lambda \mathbf{v}_n \end{bmatrix} \mathbf{p}_j - \mathbf{x}_j \right\|^2 = \sum_{j=1}^N \left\| \begin{bmatrix} \tilde{\mathbf{V}} \\ \lambda \mathbf{v}_n \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}} \\ \lambda \mathbf{v}_n \end{bmatrix}^T \mathbf{x}_j - \mathbf{x}_j \right\|^2 = \sum_{j=1}^N \left\| \begin{bmatrix} \tilde{\mathbf{V}}\tilde{\mathbf{V}}^T & \lambda \tilde{\mathbf{V}}\mathbf{v}_n \\ \lambda \mathbf{v}_n^T \tilde{\mathbf{V}}^T & \lambda^2 \mathbf{v}_n^T \mathbf{v}_n \end{bmatrix} \mathbf{x}_j - \mathbf{x}_j \right\|^2 \\
&= \sum_{j=1}^N \left\| \left(\begin{bmatrix} \tilde{\mathbf{V}}\tilde{\mathbf{V}}^T & \lambda \tilde{\mathbf{V}}\mathbf{v}_n \\ \lambda \mathbf{v}_n^T \tilde{\mathbf{V}}^T & \lambda^2 \mathbf{v}_n^T \mathbf{v}_n \end{bmatrix} - \mathbf{I} \right) \mathbf{x}_j \right\|^2 = \sum_{j=1}^N \left\| \begin{bmatrix} \tilde{\mathbf{V}}\tilde{\mathbf{V}}^T - \mathbf{I} & \lambda \tilde{\mathbf{V}}\mathbf{v}_n \\ \lambda \mathbf{v}_n^T \tilde{\mathbf{V}}^T & \lambda^2 \mathbf{v}_n^T \mathbf{v}_n - 1 \end{bmatrix} \mathbf{x}_j \right\|^2 = \sum_{j=1}^N \mathbf{x}_j^T \begin{bmatrix} \tilde{\mathbf{V}}\tilde{\mathbf{V}}^T - \mathbf{I} & \lambda \tilde{\mathbf{V}}\mathbf{v}_n \\ \lambda \mathbf{v}_n^T \tilde{\mathbf{V}}^T & \lambda^2 \mathbf{v}_n^T \mathbf{v}_n - 1 \end{bmatrix}^2 \mathbf{x}_j \\
&= \sum_{j=1}^N \begin{bmatrix} \tilde{\mathbf{x}}_j^T & x_{j,n} \end{bmatrix} \begin{bmatrix} (\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T - \mathbf{I})(\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T - \mathbf{I}) + \lambda^2 \tilde{\mathbf{V}}\mathbf{v}_n \mathbf{v}_n^T \tilde{\mathbf{V}}^T & \lambda (\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T - \mathbf{I})\tilde{\mathbf{V}}\mathbf{v}_n + \lambda^3 \tilde{\mathbf{V}}\mathbf{v}_n \mathbf{v}_n^T \tilde{\mathbf{V}}^T - \lambda \tilde{\mathbf{V}}\mathbf{v}_n \\ \lambda \mathbf{v}_n^T \tilde{\mathbf{V}}^T (\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T - \mathbf{I}) + \lambda^3 \mathbf{v}_n^T \mathbf{v}_n \mathbf{v}_n^T \tilde{\mathbf{V}}^T - \lambda \mathbf{v}_n^T \tilde{\mathbf{V}}^T & \lambda^2 \mathbf{v}_n^T \tilde{\mathbf{V}}^T \tilde{\mathbf{V}}\mathbf{v}_n + \lambda^4 (\mathbf{v}_n^T \mathbf{v}_n)^2 - 2\lambda^2 \mathbf{v}_n^T \mathbf{v}_n + 1 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_j \\ x_{j,n} \end{bmatrix} \\
&= \sum_{j=1}^N \left(\tilde{\mathbf{x}}_j^T (\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T - \mathbf{I})(\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T - \mathbf{I})\tilde{\mathbf{x}}_j + \lambda^2 \tilde{\mathbf{x}}_j^T \tilde{\mathbf{V}}\mathbf{v}_n \mathbf{v}_n^T \tilde{\mathbf{V}}^T \tilde{\mathbf{x}}_j + \lambda \tilde{\mathbf{x}}_j^T (\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T - \mathbf{I})\tilde{\mathbf{V}}\mathbf{v}_n x_{j,n} + \lambda^3 \tilde{\mathbf{x}}_j^T \tilde{\mathbf{V}}\mathbf{v}_n \mathbf{v}_n^T \tilde{\mathbf{V}}^T x_{j,n} - \lambda \tilde{\mathbf{x}}_j^T \tilde{\mathbf{V}}\mathbf{v}_n \right. \\
&\quad \left. + \lambda x_{j,n} \mathbf{v}_n^T \tilde{\mathbf{V}}^T (\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T - \mathbf{I})\tilde{\mathbf{x}}_j + \lambda^3 x_{j,n} \mathbf{v}_n^T \mathbf{v}_n \mathbf{v}_n^T \tilde{\mathbf{V}}^T \tilde{\mathbf{x}}_j - \lambda x_{j,n} \mathbf{v}_n^T \tilde{\mathbf{V}}^T \tilde{\mathbf{x}}_j + \lambda^2 x_{j,n} \mathbf{v}_n^T \tilde{\mathbf{V}}^T \tilde{\mathbf{V}}\mathbf{v}_n x_{j,n} + \lambda^4 (\mathbf{v}_n^T \mathbf{v}_n x_{j,n})^2 - 2\lambda^2 x_{j,n} \mathbf{v}_n^T \mathbf{v}_n x_{j,n} + x_{j,n}^2 \right) \\
&= \sum_{j=1}^N \left(\lambda^4 (\mathbf{v}_n^T \mathbf{v}_n x_{j,n})^2 + 2\lambda^3 \tilde{\mathbf{x}}_j^T \tilde{\mathbf{V}}\mathbf{v}_n \mathbf{v}_n^T \tilde{\mathbf{V}}^T x_{j,n} + \lambda^2 (\tilde{\mathbf{x}}_j^T \tilde{\mathbf{V}}\mathbf{v}_n \mathbf{v}_n^T \tilde{\mathbf{V}}^T \tilde{\mathbf{x}}_j + (\mathbf{v}_n^T \tilde{\mathbf{V}}^T \tilde{\mathbf{V}}\mathbf{v}_n - 2\mathbf{v}_n^T \mathbf{v}_n) x_{j,n}^2) \right. \\
&\quad \left. + 2\lambda \tilde{\mathbf{x}}_j^T (\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T - 2\mathbf{I})\tilde{\mathbf{V}}\mathbf{v}_n x_{j,n} + \tilde{\mathbf{x}}_j^T (\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T - \mathbf{I})(\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T - \mathbf{I})\tilde{\mathbf{x}}_j + x_{j,n}^2 \right).
\end{aligned}$$

Differentiating the polynomial yields:

$$P'(\lambda) = 4\lambda^3 \sum_{j=1}^N (\mathbf{v}_n^T \mathbf{v}_n x_{j,n})^2 + 6\lambda^2 \sum_{j=1}^N \tilde{\mathbf{x}}_j^T \tilde{\mathbf{V}}\mathbf{v}_n \mathbf{v}_n^T \tilde{\mathbf{V}}^T x_{j,n} + 2\lambda \sum_{j=1}^N (\tilde{\mathbf{x}}_j^T \tilde{\mathbf{V}}\mathbf{v}_n \mathbf{v}_n^T \tilde{\mathbf{V}}^T \tilde{\mathbf{x}}_j + (\mathbf{v}_n^T \tilde{\mathbf{V}}^T \tilde{\mathbf{V}}\mathbf{v}_n - 2\mathbf{v}_n^T \mathbf{v}_n) x_{j,n}^2) + 2 \sum_{j=1}^N \tilde{\mathbf{x}}_j^T (\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T - 2\mathbf{I})\tilde{\mathbf{V}}\mathbf{v}_n x_{j,n}.$$

Table 2: Derivation of the solution to (20) for SC.

Firstly, the objective function in (22) for SC can be rewritten as follows:

$$f_{SC}(\mathbf{v}_n) = \sum_{j=1}^N \left\| \begin{bmatrix} \tilde{\mathbf{V}} \\ \mathbf{v}_n^T \end{bmatrix} \mathbf{p}_j - \mathbf{x}_j \right\|^2 = \sum_{j=1}^N \left\| \begin{bmatrix} \tilde{\mathbf{V}} \\ \mathbf{v}_n^T \end{bmatrix} [\tilde{\mathbf{V}}^T \mathbf{v}_n] \mathbf{x}_j - \mathbf{x}_j \right\|^2 = \left\| \begin{bmatrix} \tilde{\mathbf{V}} \\ \mathbf{v}_n^T \end{bmatrix} [\tilde{\mathbf{V}}^T \mathbf{v}_n] \mathbf{X}^T - \mathbf{X}^T \right\|_F^2 = \left\| \left(\begin{bmatrix} \tilde{\mathbf{V}} \\ \mathbf{v}_n^T \end{bmatrix} [\tilde{\mathbf{V}}^T \mathbf{v}_n] - \mathbf{I} \right) \mathbf{X}^T \right\|_F^2,$$

where \mathbf{I} is the $n \times n$ identity matrix. Furthermore, we can express $[\tilde{\mathbf{V}}^T \mathbf{v}_n]$ as follows:

$$[\tilde{\mathbf{V}}^T \mathbf{v}_n] = \mathbf{V}^T \Delta + \mathbf{v}_n \zeta^T, \quad (28)$$

where Δ is an $n \times n$ diagonal matrix whose entries are all 1, except its n -th component, which is 0. Also, ζ is a $n \times 1$ vector whose components are all 0, except its n -th entry, which is 1. We will use (28) to rewrite $f_{SC}(\mathbf{v}_n)$ as follows:

$$f_{SC}(\mathbf{v}_n) = \left\| [(\Delta \mathbf{V} + \zeta \mathbf{v}_n^T)(\mathbf{V}^T \Delta + \mathbf{v}_n \zeta^T) - \mathbf{I}] \mathbf{X}^T \right\|_F^2 = \left\| \underbrace{(\Delta \mathbf{V} \mathbf{V}^T \Delta + \Delta \mathbf{V} \mathbf{v}_n \zeta^T + \zeta \mathbf{v}_n^T \mathbf{V}^T \Delta + \zeta \mathbf{v}_n^T \mathbf{v}_n \zeta^T) - \mathbf{I}}_{\mathbf{E}} \mathbf{X}^T \right\|_F^2.$$

Expressing the Frobenius norm as a trace yields (note that matrix the $n \times n$ matrix \mathbf{E} is symmetric):

$$f_{SC}(\mathbf{v}_n) = \text{tr}[\mathbf{X}(\mathbf{E} - \mathbf{I})^2 \mathbf{X}^T] = \text{tr}[\mathbf{X}(\mathbf{E}^2 - 2\mathbf{E} + \mathbf{I})\mathbf{X}^T] = \text{tr}[\mathbf{X}\mathbf{E}^2 \mathbf{X}^T] - 2\text{tr}[\mathbf{X}\mathbf{E}\mathbf{X}^T] + \text{tr}[\mathbf{X}\mathbf{X}^T]. \quad (29)$$

The last term in (29) does not depend on \mathbf{v}_n and is therefore irrelevant for the gradient. Thus, we will proceed by expanding the first two terms, using the following identities:

$$\Delta^2 = \Delta, \quad \zeta^T \zeta = 1, \quad \Delta \cdot \zeta = \mathbf{0}, \quad \mathbf{V}^T \Delta \mathbf{V} = \tilde{\mathbf{V}}^T \tilde{\mathbf{V}}, \quad \mathbf{X} \Delta \mathbf{V} = \tilde{\mathbf{X}} \tilde{\mathbf{V}}, \quad \text{and} \quad \mathbf{X} \zeta = \mathbf{x}_n,$$

where $\tilde{\mathbf{X}}$ is the matrix composed of the first $n - 1$ columns of \mathbf{X} , and \mathbf{x}_n is the n -th column of \mathbf{X} .

Firstly,

$$\begin{aligned} -2\text{tr}[\mathbf{X}\mathbf{E}\mathbf{X}^T] &= -2\text{tr}[\mathbf{X}\Delta\mathbf{V}\mathbf{V}^T\Delta\mathbf{X}^T] - 4\text{tr}[\mathbf{X}\zeta\mathbf{v}_n^T\mathbf{V}^T\Delta\mathbf{X}^T] - 2\text{tr}[\mathbf{X}\zeta\mathbf{v}_n\mathbf{v}_n^T\zeta^T\mathbf{X}^T] = -2\text{tr}[\tilde{\mathbf{X}}\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{X}}^T] - 4\text{tr}[\mathbf{x}_n\mathbf{v}_n^T\tilde{\mathbf{V}}^T\tilde{\mathbf{X}}^T] - 2\text{tr}[\mathbf{x}_n\mathbf{v}_n\mathbf{v}_n^T\mathbf{x}_n^T] \\ &= -2\text{tr}[\tilde{\mathbf{X}}\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{X}}^T] - 4\text{tr}[\mathbf{v}_n^T\tilde{\mathbf{V}}^T\tilde{\mathbf{X}}^T\mathbf{x}_n] - 2\text{tr}[\mathbf{x}_n^T\mathbf{x}_n\mathbf{v}_n\mathbf{v}_n^T] = -2\text{tr}[\tilde{\mathbf{X}}\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{X}}^T] - 4\mathbf{v}_n^T\mathbf{a} - 2\mathbf{x}_n^T\mathbf{x}_n\mathbf{v}_n^T\mathbf{v}_n, \end{aligned} \quad (30)$$

where $\mathbf{a} = \tilde{\mathbf{V}}^T \tilde{\mathbf{X}}^T \mathbf{x}_n$. Also, note that the first term does not depend on \mathbf{v}_n and is therefore irrelevant for the gradient.

Secondly, we proceed by expanding $\text{tr}[\mathbf{X}\mathbf{E}^2 \mathbf{X}^T]$. Since \mathbf{E} has four terms, \mathbf{E}^2 has 16, but eight of them cancel due to $\Delta \cdot \zeta = \mathbf{0}$. Also, some terms appear twice. In particular, we have:

$$\begin{aligned} \text{tr}[\mathbf{X}\mathbf{E}^2 \mathbf{X}^T] &= \text{tr}[\mathbf{X}\Delta\mathbf{V}\mathbf{V}^T\Delta^2\mathbf{V}\mathbf{V}^T\Delta\mathbf{X}^T] + 2\text{tr}[\mathbf{X}\zeta\mathbf{v}_n^T\mathbf{V}^T\Delta^2\mathbf{V}\mathbf{V}^T\Delta\mathbf{X}^T] + \text{tr}[\mathbf{X}\Delta\mathbf{V}\mathbf{v}_n\zeta^T\zeta\mathbf{v}_n^T\mathbf{V}^T\Delta\mathbf{X}^T] \\ &\quad + \text{tr}[\zeta\mathbf{v}_n^T\mathbf{V}^T\Delta^2\mathbf{V}\mathbf{v}_n\zeta^T] + 2\text{tr}[\mathbf{X}\zeta\mathbf{v}_n^T\mathbf{V}^T\zeta\mathbf{v}_n^T\mathbf{V}^T\Delta\mathbf{X}^T] + \text{tr}[\mathbf{X}\zeta\mathbf{v}_n\mathbf{v}_n^T\zeta^T\zeta\mathbf{v}_n\mathbf{v}_n^T\zeta^T\mathbf{X}^T] \\ &= \text{tr}[\tilde{\mathbf{X}}\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{X}}^T] + 2\text{tr}[\mathbf{x}_n\mathbf{v}_n^T\tilde{\mathbf{V}}^T\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{X}}^T] + \text{tr}[\tilde{\mathbf{X}}\tilde{\mathbf{V}}\mathbf{v}_n\mathbf{v}_n^T\tilde{\mathbf{V}}^T\tilde{\mathbf{X}}^T] + \text{tr}[\mathbf{x}_n\mathbf{v}_n^T\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\mathbf{v}_n\mathbf{x}_n^T] + 2\text{tr}[\mathbf{x}_n\mathbf{v}_n^T\mathbf{v}_n\mathbf{v}_n^T\tilde{\mathbf{V}}^T\tilde{\mathbf{X}}^T] + \text{tr}[\mathbf{x}_n\mathbf{v}_n^T\mathbf{v}_n\mathbf{v}_n^T\mathbf{x}_n^T] \\ &= \text{tr}[\tilde{\mathbf{X}}\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{X}}^T] + 2\text{tr}[\mathbf{v}_n^T\tilde{\mathbf{V}}^T\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{X}}^T\mathbf{x}_n] + \text{tr}[\mathbf{v}_n^T\tilde{\mathbf{V}}^T\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\tilde{\mathbf{V}}\mathbf{v}_n] + \mathbf{x}_n^T\mathbf{x}_n\text{tr}[\mathbf{v}_n^T\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\mathbf{v}_n] + 2\text{tr}[\mathbf{v}_n^T\mathbf{v}_n\mathbf{v}_n^T\tilde{\mathbf{V}}^T\tilde{\mathbf{X}}^T\mathbf{x}_n] + \mathbf{x}_n^T\mathbf{x}_n\text{tr}[\mathbf{v}_n^T\mathbf{v}_n\mathbf{v}_n^T\mathbf{v}_n] \\ &= \text{tr}[\tilde{\mathbf{X}}\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{X}}^T] + 2\text{tr}[\mathbf{v}_n^T\mathbf{D}\mathbf{a}] + \text{tr}[\mathbf{v}_n^T\mathbf{C}\mathbf{v}_n] + \mathbf{x}_n^T\mathbf{x}_n\text{tr}[\mathbf{v}_n^T\mathbf{D}\mathbf{v}_n] + 2\text{tr}[\mathbf{v}_n^T\mathbf{v}_n\mathbf{v}_n^T\mathbf{a}] + \mathbf{x}_n^T\mathbf{x}_n\text{tr}[\mathbf{v}_n^T\mathbf{v}_n\mathbf{v}_n^T\mathbf{v}_n] \\ &= \text{tr}[\tilde{\mathbf{X}}\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{X}}^T] + 2(\mathbf{v}_n^T\mathbf{D}\mathbf{a}) + \mathbf{v}_n^T\mathbf{C}\mathbf{v}_n + \mathbf{x}_n^T\mathbf{x}_n(\mathbf{v}_n^T\mathbf{D}\mathbf{v}_n) + 2(\mathbf{v}_n^T\mathbf{v}_n\mathbf{v}_n^T\mathbf{a}) + \mathbf{x}_n^T\mathbf{x}_n(\mathbf{v}_n^T\mathbf{v}_n)(\mathbf{v}_n^T\mathbf{v}_n), \end{aligned} \quad (31)$$

where $\mathbf{D} = \tilde{\mathbf{V}}^T \tilde{\mathbf{V}}$, and $\mathbf{C} = \tilde{\mathbf{V}}^T \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} \tilde{\mathbf{V}}$, which are both symmetric.

Substituting (30) and (31) in (29) we have:

$$\begin{aligned} f_{SC}(\mathbf{v}_n) &= -4\mathbf{v}_n^T\mathbf{a} - 2\mathbf{x}_n^T\mathbf{x}_n\mathbf{v}_n^T\mathbf{v}_n + 2\mathbf{v}_n^T\mathbf{D}\mathbf{a} + \mathbf{v}_n^T\mathbf{C}\mathbf{v}_n + \mathbf{x}_n^T\mathbf{x}_n\mathbf{v}_n^T\mathbf{D}\mathbf{v}_n + 2\mathbf{v}_n^T\mathbf{v}_n\mathbf{v}_n^T\mathbf{a} + \mathbf{x}_n^T\mathbf{x}_n\mathbf{v}_n^T\mathbf{v}_n\mathbf{v}_n^T\mathbf{v}_n \\ &\quad + \text{tr}[\mathbf{X}\mathbf{X}^T] - 2\text{tr}[\tilde{\mathbf{X}}\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{X}}^T] + \text{tr}[\tilde{\mathbf{X}}\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{X}}^T], \end{aligned}$$

where the terms involving traces do not depend on \mathbf{v}_n and are therefore irrelevant for computing the gradient of the function.

Finally, taking the derivative with respect to \mathbf{v}_n yields:

$$\begin{aligned} \nabla f_{SC}(\mathbf{v}_n) &= -4\mathbf{a} - 4\mathbf{x}_n^T\mathbf{x}_n\mathbf{v}_n + 2\mathbf{D}\mathbf{a} + 2\mathbf{C}\mathbf{v}_n + 2\mathbf{x}_n^T\mathbf{x}_n\mathbf{D}\mathbf{v}_n + 4\mathbf{v}_n\mathbf{v}_n^T\mathbf{a} + 2\mathbf{v}_n^T\mathbf{v}_n\mathbf{a} + 4\mathbf{x}_n^T\mathbf{x}_n\mathbf{v}_n\mathbf{v}_n^T\mathbf{v}_n \\ &= 2\mathbf{C}\mathbf{v}_n + 2\mathbf{v}_n^T\mathbf{v}_n\mathbf{a} + (2\mathbf{D} - 4\mathbf{I} + 4\mathbf{v}_n\mathbf{v}_n^T)(\mathbf{a} + \mathbf{x}_n^T\mathbf{x}_n\mathbf{v}_n), \end{aligned}$$

where we have used the following rules:

$$\frac{\partial \mathbf{x}^T \mathbf{b}}{\partial \mathbf{x}} = \mathbf{b}, \quad \frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^T) \mathbf{x}, \quad \frac{\partial \mathbf{x}^T \mathbf{x} \mathbf{x}^T \mathbf{b}}{\partial \mathbf{x}} = 4\mathbf{x} \mathbf{x}^T \mathbf{b} + 2\mathbf{x}^T \mathbf{x} \mathbf{b}, \quad \text{and} \quad \frac{\partial \mathbf{x}^T \mathbf{x} \mathbf{x}^T \mathbf{x}}{\partial \mathbf{x}} = 4\mathbf{x} \mathbf{x}^T \mathbf{x},$$

where $\mathbf{x}, \mathbf{b} \in \mathbb{R}^n$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$.

Table 3: Derivation of the gradient of the objective function in (22) for SC.