SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

Robust l_1 Approaches to Computing the Geometric Median and Principal and Independent Components

S.L. Keeling K. Kunisch

SFB-Report No. 2014-003 March 2014

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

Der Wissenschaftsfonds.

SFB sponsors:

- **Austrian Science Fund** (FWF)
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**

Robust ℓ_1 Approaches to Computing the Geometric Median and Principal and Independent Components

Stephen L. Keeling[1] and Karl Kunisch[1]

Abstract. Robust measures are introduced for methods to determine statistically uncorrelated or also statistically independent components spanning data measured in a way that does not permit direct separation of these underlying components. Because of the nonlinear nature of the proposed methods, iterative primal-dual based methods are presented for the optimization of merit functions, and convergence of these methods is proved. Illustrative examples are presented to demonstrate the benefits of the robust approaches, including an application to the processing of dynamic medical imaging.

Keywords: geometric median, principal component analysis, independent component analysis, robustness, convergence of iterative methods

1 Introduction

The topics of this work focus on the low-dimensional representation of complex measured data. The lowest dimensional representation is a type of average. More accurate representations add dimensions beyond the average based upon subspaces in which the data vary the most. Choosing a basis for such subspaces is driven by the priority that data coordinates with respect to this basis be statistically uncorrelated or even statistically independent. The particular interest here is to present methods for performing these tasks which are robust against outliers in the measured data.

The most common type of average is the mean, which may be formulated variationally as the point minimizing the sum of squared distances to data points. As discussed in [10], a more robust method involves minimizing a merit function which does not grow as rapidly with respect to the data and would thereby apply less weight to erroneous data points far from a natural average. Various notions of an average based upon ℓ_1 measures are discussed in [8]. Based upon examples presented in Section 3, the type of average selected for this work is the *geometric median*, which may be formulated variationally as the point minimizing the sum of distances (not squared) to data points. The problem of determining the geometric median has a long history. In the 1937 paper by Weiszfeld [15], three proofs concerning the uniqueness of the geometric median are given, and one of these supplies an algorithm for its computation. We also refer to the recent annotated translation of that paper [13]. See also [4]. A shorter proof of uniqueness is given in Section 6 which is based upon a strict convexity argument. Moreover, a possibly novel characterization of a solution is provided in case data points are colinear. A primal-dual iteration for computing the geometric median is proposed in Section 3, and the convergence of this scheme is proved in Section 6. An alternative approach based upon the Chambolle-Pock algorithm [6] is presented in the appendix, and its performance is compared to our approach.

Given a natural average or center of the measured data, one may then wish to determine the direction in which data points vary the most from the center. This direction is the most significant *principal component* of the data. Principal components of lesser significance are sought similarly but within the orthogonal complement to the more significant ones. Determining and analysing such components is the subject of principal component analysis (PCA) [11]. The most common way of determining these components is to select them as the eigenvectors of the covariance matrix of the data. The more significant components correspond to the larger eigenvalues of the covariance matrix since each eigenvalue gives the variance of the data projected onto the corresponding eigenvector. As discussed in Section 4, determining each eigenvector can be formulated variationally in terms of finding a best fit line through

[1]Institut für Mathematik und Wissenschaftliches Rechnen, Karl-Franzens-Universität Graz, Heinrichstraße 36, 8010 Graz, Austria; email: firstname.lastname@uni-graz.at. The authors are supported by the Austrian Science Fund *Fonds zur Förderung der Wissenschaftlichen Forschung* (FWF) under grant SFB F032 and are linked on the SFB webpage http://math.uni-graz.at/mobis/.

the data, where the line minimizes the sum of squared distances to data points. See [3] and [7] for ℓ_1 based alternatives to this criterion. Based upon examples presented in Section 4, this line is determined here more robustly by minimizing the sum of distances (not squared) to data points. In other words, analogous to defining an average as a geometric median *point*, a principal component is defined as a geometric median *line*. In Section 4 an iterative scheme is proposed for computing this line. However, since the merit function is not convex, uniqueness of minimizers cannot be expected. Nevertheless, convergence of the scheme to a minimizer is proved in Section 7. This scheme, just as the scheme used for computing the geometric median, is based on a primal-dual formulation of the optimality condition. For a description of a greedy-algorithm to an ℓ_1 maximization approach to PCA we refer to [12].

Suppose that the data are rotated to an axis system aligned with principal components and that they are then scaled along each new axis to normalize the respective variances to unity. When this rotation and scaling is carried out by standard methods using ℓ_2 measures, the transformed data have a covariance matrix equal to the identity. Then the data are said to have been *sphered*. In particular, the new data coordinates are statistically uncorrelated. However, they are not necessarily statistically independent [10]. (See, e.g., the example of Fig. 3 with $m = 1$ in (5.1) so that the data are sphered but the coordinates do not satisfy the independence criterion (5.4).) It might then be postulated that the data can be represented in a rotated axis system with respect to which coordinates are statistically independent. Determining and analysing such a system is the subject of independent component analysis (ICA). In case the postulate holds, coordinates of the sphered data represent weighted sums of statistically independent variables, and by the Central Limit Theorem [1] histograms of such coordinates tend to be bell shaped. In order to identify the postulated rotation, it is standard to minimize the *Gaussianity* of histograms of coordinates in the desired rotated system. The approach proposed by [10] is to determine this rotation by maximizing a merit function which is known to be minimized by data with a Gaussian distribution. It is also argued in [10] that one such merit function is more robust to data outliers than another when it does not grow as rapidly with respect to the data. Such candidate merit functions are considered in Section 5. The optimization method of [17] is robust against local extrema. The approach proposed here for determining the desired rotation begins by targeting independence directly instead of using the indirect measure of Gaussianity. The merit function proposed in Section 5 is motivated by the observation that while sphered axes tend to be aligned with data clusters, independent axes tend to *separate* clusters. See the examples presented in Section 5 for details. A fixed point iteration scheme based on the primal-dual formulation of the optimality condition is proposed in Section 5 for computing robust independent components, and the convergence of this scheme is proved in Section 8.

The paper is outlined as follows. In Section 2, standard ℓ_2 approaches to PCA and ICA are summarized, particularly to establish the background used later for the presentation of more robust methods. In Section 3 a robust method of data centering is proposed using the geometric median. In Section 4 a robust method for determining principal components is proposed using lines which are best fit in the sense that the sum of distances (not squared) to the data points is minimized within the subspace orthogonal to other components. In Section 5 a robust method for determining independent components is proposed which maximizes separations among sphered data clusters. Due to the nonlinearity of the respective optimality conditions, iterative schemes are proposed in Sections 3 – 5 to solve the respective optimization problems. Convergence of these schemes is proved in Sections 6 – 8. In Section 9 the proposed methods are applied to a magnetic resonance image sequence to separate intensity changes due to physiological motion from those due to contrast agent, and benefits of the robust methods are demonstrated with respect to this realistic example. See also [14] and [16]. The paper ends with a summary in Section 10.

2 Summary of ℓ_2 Approaches to PCA and ICA

Let an unknown random vector $\mathrm{z} \in \mathbb{R}^m$ be given with components $\{\mathrm{z}_i\}_{i=1}^m$ which will be called *sources*. For example, the sources could be random variables associated with sounds produced independently at a *cocktail party*. The sources are assumed to satisfy the following:

1. For $1 \le i \ne j \le m$, z_i and z_j are statistically independent.

2. No z_i is normally distributed.

3. For $1 \leq i \leq m$, the variance $\sigma_i^2 = \mathrm{E}[(\mathrm{z}_i - \mathrm{E}[\mathrm{z}_i])^2]$ of z_i is positive.

Here, E denotes the expectation. Since the sources are statistically independent, they are uncorrelated [10]. Let their positive definite diagonal covariance matrix be denoted by

$$C(\mathrm{z}) = \{\mathrm{E}[\mathrm{z}_i - \mathrm{E}[\mathrm{z}_i], \mathrm{z}_j - \mathrm{E}[\mathrm{z}_j]]\}_{i,j=1}^m = \mathrm{diag}\{\sigma_i^2\}_{i=1}^m$$

which is unknown. Let a random vector $\mathrm{y} \in \mathbb{R}^m$ be defined through a measurement process

$$\mathrm{y} = A\mathrm{z} \tag{2.1}$$

modelled in terms of the *mixing matrix* $A \in \mathbb{R}^{m\times m}$. The components $\{\mathrm{y}_i\}_{i=1}^m$ of y will be called measurements. For example, the measurements could be random variables associated with sounds recorded by separate microphones at the cocktail party mentioned above. Under the assumption that the mixing matrix is invertible, the goal is to determine a matrix $W \in \mathbb{R}^{m\times m}$ such that the components $\{\mathrm{x}_i\}_{i=1}^m$ of the random vector

$$\mathrm{x} = W\mathrm{y} \tag{2.2}$$

estimate the sources in the following sense. First, normalizing $\mathrm{z} = A^{-1}\mathrm{y}$ according to $C(\mathrm{z})^{-\frac{1}{2}}\mathrm{z}$ removes the ambiguity of unknown variances by setting the covariance matrix to the identity. Secondly, since the order and sign of components in $C(\mathrm{z})^{-\frac{1}{2}}\mathrm{z}$ is unknown, the alternative $PC(\mathrm{z})^{-\frac{1}{2}}\mathrm{z}$ also satisfies the source assumptions when $P \in \mathbb{R}^{m\times m}$ is any matrix satisfying $(P_{q_i,j})^2 = \delta_{i,j}$ with $\{q_i\}_{i=1}^m$ begin a permutation of $\{i\}_{i=1}^m$. Thus, W estimates a product $PC(\mathrm{z})^{-\frac{1}{2}}A^{-1}$, and the covariance matrix of x in (2.2) is the identity.

Suppose that each random measurement variable y_i is sampled directly to obtain n samples $\{y_{ij}\}_{j=1}^n$. Implicitly underlying these are samples $\{z_{ij}\}_{j=1}^n$ of each random source variable z_i. Define the sample vectors $\boldsymbol{y}_i = \{y_{ij}\}_{j=1}^n$, $\boldsymbol{z}_i = \{z_{ij}\}_{j=1}^n$, $i = 1, \ldots, m$. According to the linear model in (2.1), the matrices $Y = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m\}^{\mathrm{T}} \in \mathbb{R}^{m\times n}$ and $Z = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m\}^{\mathrm{T}} \in \mathbb{R}^{m\times n}$ are related by

$$Y = AZ \tag{2.3}$$

By (2.2) the estimation $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}^{\mathrm{T}} \in \mathbb{R}^{m\times n}$ of the sources satisfies

$$X = WY \tag{2.4}$$

The matrix W is determined stepwise in terms of its singular value decomposition

$$W = U\Lambda^{-\frac{1}{2}}V^{\mathrm{T}} \tag{2.5}$$

where $U, V \in \mathbb{R}^{m\times m}$ are orthogonal and $\Lambda \in \mathbb{R}^{m\times m}$ is positive definite and diagonal. Specifically, after the data are centered

$$Y_{\mathrm{c}} = Y - \bar{Y} \tag{2.6}$$

with

$$\bar{Y} = \{\bar{\boldsymbol{y}}_1, \ldots, \bar{\boldsymbol{y}}_m\}^{\mathrm{T}}, \quad \bar{\boldsymbol{y}}_i = \frac{1}{n}\sum_{j=1}^n y_{ij} \tag{2.7}$$

the product $V^{\mathrm{T}}Y_{\mathrm{c}}$ should rotate the data so that the new coordinate axes are aligned with the visually natural axes of the cluster of data points $\{Y\hat{\boldsymbol{e}}_j\}_{j=1}^n$, $\hat{\boldsymbol{e}}_j \in \mathbb{R}^n$, $(\hat{\boldsymbol{e}}_j)_i = \delta_{i,j}$. After this rotation, the product

$$Y_{\mathrm{s}} = \Lambda^{-\frac{1}{2}}V^{\mathrm{T}}Y_{\mathrm{c}} \tag{2.8}$$

should scale the data so that the variance along each new coordinate axis is unity. For this reason, the data Y_{s} are said to be *sphered*. The final orthogonal matrix U in (2.5) is chosen so that the components of the random variable x in (2.2) are maximally independent in a sense made precise below.

To determine the transformations V and Λ, the covariance matrix of the sphered data is required to be the identity,

$$I = \tfrac{1}{n} Y_{\mathrm{s}} Y_{\mathrm{s}}^{\mathrm{T}} = \Lambda^{-\frac{1}{2}} V^{\mathrm{T}} [\tfrac{1}{n} Y_{\mathrm{c}} Y_{\mathrm{c}}^{\mathrm{T}}] V \Lambda^{-\frac{1}{2}} \tag{2.9}$$

which is accomplished by determining the matrices V and Λ from the eigenspace decomposition of the centered data,

$$\tfrac{1}{n} Y_{\mathrm{c}} Y_{\mathrm{c}}^{\mathrm{T}} = V \Lambda V^{\mathrm{T}}, \quad V^{\mathrm{T}} V = I. \tag{2.10}$$

The columns of V are the so-called *principal components* of the data Y. Analyzing this decomposition is the subject of *principal component analysis* (PCA). For instance, the sampled data may be filtered by projecting these data onto subspaces spanned by principal components. For this, assume that the entries of $\Lambda = \mathrm{diag}\{\lambda_i\}_{i=1}^m$ and $V = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ are ordered according to $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. This means that the variance $\lambda_i = \frac{1}{n} \|Y_{\mathrm{c}}^{\mathrm{T}} \boldsymbol{v}_i\|_{\ell_2}^2$ of the data Y_{c} along the axis $\boldsymbol{v}_i$ is larger than the variance λ_j along the axis $\boldsymbol{v}_j$ for $i < j$. To select only the $r < m$ components with respect to which the data have the most variation, define the projected data $Y_P \approx Y$ by

$$Y_P = \overline{Y} + V \Lambda^{\frac{1}{2}} P^{\mathrm{T}} P \Lambda^{-\frac{1}{2}} V^{\mathrm{T}} (Y - \overline{Y}) \tag{2.11}$$

where the projector $P \in \mathbb{R}^{r \times m}$ is defined with entries $P_{i,j} = \delta_{i,j}$. Note that with (2.6), (2.8) and (2.10), this result can be rewritten as $Y_P = \overline{Y} + \frac{1}{n} Y_{\mathrm{c}} (P Y_{\mathrm{s}})^{\mathrm{T}} (P Y_{\mathrm{s}})$.

Next, the transformation U in (2.1) is determined so that the components of the random variable x in (2.2) are independent. While the rows of the sphered data Y_{s} are statistically uncorrelated, they are not necessarily statistically independent [10]. A criterion is now sought for a final rotation of axes which gives the desired independence. Since as seen in (2.1) measurements are sums of independent random variables, the Central Limit Theorem suggests why the measurements tend to be normally distributed [1]. The matrix U is often chosen to reverse this effect, i.e., to make the components of x depart from being normally distributed as much as possible. Here, the significance of the assumption that no component z_i be normally distributed can be seen, as otherwise the proposed measure of independence would not bring a separation of sources in the following. For the required statistical constructions, let $\mathrm{E}[\mathrm{x}]$ denote the expectation of a random variable x. Since a normally distributed random variable n with mean 0 and variance σ^2 has moments

$$\mathrm{E}[|\mathrm{n}|^m] = \kappa_m \sigma^m, \quad \kappa_m = \begin{cases} (m-1)!!, & m \text{ even} \\ \sqrt{\frac{2}{\pi}} (m-1)!!, & m \text{ odd} \end{cases} \tag{2.12}$$

it follows that the Kurtosis $K = K_4$,

$$K_m(\mathrm{x}) = \mathrm{E}[|\mathrm{x} - \mathrm{E}[\mathrm{x}]|^m] - \kappa_m \mathrm{E}[|\mathrm{x} - \mathrm{E}[\mathrm{x}]|^2]^{m/2} \tag{2.13}$$

of n satisfies $K(\mathrm{n}) = \mathrm{E}[|\mathrm{n}|^4] - 3\mathrm{E}[|\mathrm{n}|^2]^2 = 0$. Hence, a parameter dependent random variable may be made to depart maximally from being normally distributed by maximizing the square of its Kurtosis with respect to parameters. Applying this criterion to the rows of

$$X_{\mathrm{c}} = U Y_{\mathrm{s}} \tag{2.14}$$

the rows of $U = \{\boldsymbol{u}_i^{\mathrm{T}}\}_{i=1}^m$ are determined as follows. Define

$$U_l = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_l\}^{\mathrm{T}}, \quad U_l \in \mathbb{R}^{l \times m}, \quad l = 1, \ldots, m, \quad U_0 = \{\} \tag{2.15}$$

and the projected data

$$Y_l = (I - U_{l-1}^{\mathrm{T}} U_{l-1}) Y_{\mathrm{s}}, \quad l = 2, \ldots, m, \quad Y_1 = Y_{\mathrm{s}} \tag{2.16}$$

whose columns lie in $T_l \subset \mathbb{R}^m$ defined as the range of Y_l. Note that

$$\boldsymbol{u}_l^{\mathrm{T}} Y_l = \boldsymbol{u}_l^{\mathrm{T}} Y_{\mathrm{s}} - \boldsymbol{u}_l^{\mathrm{T}} U_{l-1}^{\mathrm{T}} U_{l-1} Y_{\mathrm{s}} = \boldsymbol{u}_l^{\mathrm{T}} Y_{\mathrm{s}}, \quad \boldsymbol{u} \in T_l. \tag{2.17}$$

Given T_l, let the l-th column of U^{T} be determined inductively by

$$\boldsymbol{u}_l = \frac{\mathbf{u}_l}{\|\mathbf{u}_l\|_{\ell_2}}, \quad \mathbf{u}_l = \underset{\mathbf{u}\in T_l}{\operatorname{argmax}} F\left(\frac{\mathbf{u}^{\mathrm{T}} Y_l}{\|\mathbf{u}\|_{\ell_2}}\right), \quad F(\boldsymbol{x}) = K^2(\boldsymbol{x}). \tag{2.18}$$

By (2.9) and (2.17) the second moment of $\mathbf{u}^{\mathrm{T}} Y_l/\|\mathbf{u}\|_{\ell_2}$ is

$$\mathbf{u}^{\mathrm{T}}[Y_l Y_l^{\mathrm{T}}/n]\mathbf{u}/\|\mathbf{u}\|_{\ell_2}^2 = \mathbf{u}^{\mathrm{T}}[Y_{\mathrm{s}} Y_{\mathrm{s}}^{\mathrm{T}}/n]\mathbf{u}/\|\mathbf{u}\|_{\ell_2}^2 = \mathbf{u}^{\mathrm{T}}\mathbf{u}/\|\mathbf{u}\|_{\ell_2}^2 = 1, \quad \mathbf{u} \in T_l. \tag{2.19}$$

Thus, F in (2.18) is computed according to

$$F(\boldsymbol{x}) = \left[\frac{1}{n}\|\boldsymbol{x}\|_{\ell_4}^4 - 3\right]^2 \quad \text{for} \quad \frac{1}{n}\|\boldsymbol{x}\|_{\ell_2}^2 = 1. \tag{2.20}$$

In this way the rows of U are determined sequentially so that the earlier components of x depart from being normally distributed more than later components. Alternatively, all components of x may be estimated with roughly equal quality by determining all rows of U simultaneously through maximizing $\sum_{l=1}^m F(\boldsymbol{u}_l^{\mathrm{T}} Y_{\mathrm{s}})$ under the conditions $\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{u}_j = \delta_{i,j}$, $1 \le i,j \le m$. Once matrices U, Λ and V are determined, the source samples are estimated according to (2.4) and (2.5).

The columns of $V\Lambda^{\frac{1}{2}}U^{\mathrm{T}}$ are the so-called *independent components* of the data Y. Analyzing this decomposition is the subject of *independent component analysis* (ICA). For instance, the sampled data may be filtered by projecting these data onto subspaces spanned by independent components. Specifically, to select the $r < m$ desired independent components $\{q_1, \ldots, q_r\} \subset \{1, \ldots, m\}$, define the projected data $Y_Q \approx Y$ by

$$Y_Q = \overline{Y} + V\Lambda^{\frac{1}{2}}U^{\mathrm{T}}Q^{\mathrm{T}}QU\Lambda^{-\frac{1}{2}}V^{\mathrm{T}}(Y - \overline{Y}) \tag{2.21}$$

where the projector $Q \in \mathbb{R}^{r\times m}$ is defined with entries $Q_{ij} = \delta_{q_i,j}$. Note that with (2.6), (2.8), (2.10) and (2.14), this result can be rewritten as $Y_Q = \overline{Y} + \frac{1}{n}Y_{\mathrm{c}}(QX_{\mathrm{c}})^{\mathrm{T}}(QX_{\mathrm{c}})$.

In the calculations above it is implicitly assumed that the number of samples n is at least as large as the number of sources m. Otherwise, the rank n of the covariance matrix $\frac{1}{n}Y_{\mathrm{s}}Y_{\mathrm{s}}^{\mathrm{T}}$ would be less than its dimension m, and the diagonal matrix Λ in (2.10) would not be positive definite. In case $n < m$ does in fact hold, because so few samples have been collected, one might be inclined simply to replace Y with Y^{T} and thereby reverse the roles of time and space in the data. However, the data must possess an ergodicity property for the results with transposed data to be roughly equivalent to those without transposed data. Since such a property may not generally hold, the matrices above are determined here as follows; see also [5]. With $\overline{Y}$ and Y_{c} given by (2.6), define the singular value decomposition $Y_{\mathrm{c}}/\sqrt{n} = V\hat{\Sigma}\hat{Y}_{\mathrm{s}}/\sqrt{n}$ in terms of rotation matrices $V \in \mathbb{R}^{m\times m}$ and $\hat{Y}_{\mathrm{s}}/\sqrt{n} \in \mathbb{R}^{n\times n}$ and a rectangular matrix $\hat{\Sigma} \in \mathbb{R}^{m\times n}$ for which $\hat{\Lambda} = \hat{\Sigma}^{\mathrm{T}}\hat{\Sigma} \in \mathbb{R}^{n\times n}$ is diagonal and positive definite. The matrices $\hat{\Lambda}$ and $\hat{Y}_{\mathrm{s}}$ are determined from the eigenspace decomposition $Y_{\mathrm{c}}^{\mathrm{T}}Y_{\mathrm{c}} = \hat{Y}_{\mathrm{s}}^{\mathrm{T}}\hat{\Lambda}\hat{Y}_{\mathrm{s}}$, $\hat{Y}_{\mathrm{s}}\hat{Y}_{\mathrm{s}}^{\mathrm{T}} = nI$. Since the last $m-n$ rows of $\hat{\Sigma}$ are zero, the last $m-n$ columns $\tilde{V} \in \mathbb{R}^{m\times(m-n)}$ of $V = [\hat{V}, \tilde{V}]$ may be neglected to obtain $Y_{\mathrm{c}} = \hat{V}\hat{\Lambda}^{\frac{1}{2}}\hat{Y}_{\mathrm{s}}$. The matrix $\hat{V} \in \mathbb{R}^{m\times n}$ is determined from $\hat{V} = Y_{\mathrm{c}}\hat{Y}_{\mathrm{s}}^{\mathrm{T}}\hat{\Lambda}^{-\frac{1}{2}}/n$. The sphered data $\hat{Y}_{\mathrm{s}}$ are transformed by the rotation matrix $\hat{U} \in \mathbb{R}^{n\times n}$ maximizing independence of the rows of $\hat{X}_{\mathrm{c}} = \hat{U}\hat{Y}_{\mathrm{s}} \in \mathbb{R}^{n\times n}$. Note that with $\Lambda = \hat{\Sigma}\hat{\Sigma}^{\mathrm{T}}$ it also follows that $Y_{\mathrm{c}}Y_{\mathrm{c}}^{\mathrm{T}} = V\hat{\Sigma}[\hat{Y}_{\mathrm{s}}\hat{Y}_{\mathrm{s}}^{\mathrm{T}}]\hat{\Sigma}^{\mathrm{T}}V^{\mathrm{T}} = nV\Lambda V^{\mathrm{T}}$ holds, giving (2.10). Let $\hat{Y}_{\mathrm{s}}$ be padded with $m-n$ zero rows to obtain $Y_{\mathrm{s}} = [\hat{Y}_{\mathrm{s}}; 0] \in \mathbb{R}^{m\times n}$. With $\Sigma = \Lambda^{\frac{1}{2}}$ it follows from the singular value decomposition of Y_{c} that $Y_{\mathrm{c}} = V\Sigma Y_{\mathrm{s}}$ holds, and hence the counterpart $Y_{\mathrm{s}} = (\Lambda^{\dagger})^{\frac{1}{2}}V^{\mathrm{T}}Y_{\mathrm{c}}$ to (2.8) holds, where $\Lambda^{\dagger}$ denotes the pseudo-inverse of Λ. The rotation matrix $U = [[\hat{U}, 0]; [0, \tilde{U}]] \in \mathbb{R}^{m\times m}$ can be defined by supplementing $\hat{U}$ with the rotation matrix $\tilde{U} \in \mathbb{R}^{(m-n)\times(m-n)}$ and otherwise padding with zeros, and $X_{\mathrm{c}} = UY_{\mathrm{s}} \in \mathbb{R}^{m\times n}$ can be defined to give (2.14). However, according to $X_{\mathrm{c}} = UY_{\mathrm{s}} = [[\hat{U}, 0]; [0, \tilde{U}]][\hat{Y}_{\mathrm{s}}; 0] = [\hat{X}_{\mathrm{c}}; 0]$, the last $m-n$ rows of X_{c} are zero, contrary to the objective that the rows be independent. Thus, $\hat{X}_{\mathrm{c}}$ marks the end of the calculation and $\hat{X} = \hat{U}\hat{\Lambda}^{-\frac{1}{2}}\hat{V}^{\mathrm{T}}Y$ gives the maximum number of independent components which can be determined from the undersampled data. Finally, for projectors $P, Q \in \mathbb{R}^{n\times n}$, (2.11) and (2.21) become $Y_P = \overline{Y} + \frac{1}{n}Y_{\mathrm{c}}(P\hat{Y}_{\mathrm{s}})^{\mathrm{T}}(P\hat{Y}_{\mathrm{s}})$ and $Y_Q = \overline{Y} + \frac{1}{n}Y_{\mathrm{c}}(Q\hat{X}_{\mathrm{s}})^{\mathrm{T}}(Q\hat{X}_{\mathrm{s}})$, respectively.

3 ℓ_1 Approach to Centering

That ℓ_1 measures lead to statistically robust results may be highlighted by the following simple example. Suppose samples $Y = \{y_j\}_{j=1}^n = \{0, 1, \ldots, 1\} \in \mathbb{R}^n$ have been collected, where the first measurement is clearly an outlier. The ℓ_2 mean of these data is given by

$$\operatorname*{argmin}_{\mu\in\mathbb{R}} \sum_{j=1}^{n} |y_j - \mu|^2 = \text{mean}(\boldsymbol{y}) = (n-1)/n \tag{3.1}$$

which is clearly influenced by the outlier. On the other hand, the ℓ_1 mean is given by

$$\operatorname*{argmin}_{\mu\in\mathbb{R}} \sum_{j=1}^{n} |y_j - \mu| = \text{median}(\boldsymbol{y}) = 1 \tag{3.2}$$

which is insensitive to the outlier.

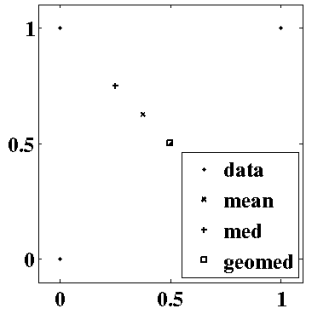
For a generalization of this robust scalar mean to its counterpart for vectors, let the data $Y = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m\}^{\mathrm{T}} \in \mathbb{R}^{m\times n}$, $\boldsymbol{y}_i = \{y_{ij}\}_{j=1}^n$, with columns $Y\hat{\boldsymbol{e}}_j$, $\hat{\boldsymbol{e}}_j \in \mathbb{R}^n$, $(\hat{\boldsymbol{e}}_j)_i = \delta_{i,j}$, be given as in Section 2. Note that if the ℓ_1 mean were defined according to

$$\operatorname*{argmin}_{\boldsymbol{\mu}\in\mathbb{R}^m} \sum_{j=1}^{n} \|Y\hat{\boldsymbol{e}}_j - \boldsymbol{\mu}\|_{\ell_1} = \left\{ \operatorname*{argmin}_{\mu_i\in\mathbb{R}} \sum_{j=1}^{n} |y_{ij} - \mu_i| \right\}_{i=1}^{m} = \left\{\text{median}\{y_{ij}\}_{j=1}^n\right\}_{i=1}^m \tag{3.3}$$

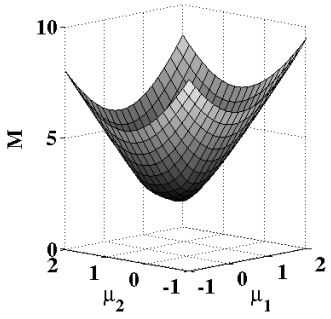
then the solution would be unnaturally determined componentwise through decoupled minimizations. By contrast, the following ℓ_1 mean for vectors [8], i.e., the geometric median [13],

$$\overline{Y} = \operatorname*{argmin}_{\boldsymbol{\mu}\in\mathbb{R}^m} M(\boldsymbol{\mu}), \quad M(\boldsymbol{\mu}) = \sum_{j=1}^{n} \|Y\hat{\boldsymbol{e}}_j - \boldsymbol{\mu}\|_{\ell_2} \tag{3.4}$$

minimizes, in a natural way, the ℓ_1 norm of Euclidean distances between the data points and the selected mean. The robustness of this measure in relation to the mean or median can be highlighted by the following simple example, which is illustrated in Fig. 1a. Here the data are given by



(a) data, mean, median and geometric median



(b) merit function $M(\mu_1, \mu_2)$

Figure 1: The mean, median and geometric median are compared in (a), where the data of (3.5) are shown with $\cdot$, the mean with $\times$, the median with $+$ and the geometric median with $\square$. Shown in (b) is the landscape of the merit function M in (3.4).

$$Y = \begin{bmatrix} 0 & \frac{1}{2} & 1 & 0 \\ 0 & \frac{1}{2} & 1 & 1 \end{bmatrix} \tag{3.5}$$

marked with $\cdot$ in Fig. 1a, and the point $(0, 1)$ may be regarded as an outlier from points otherwise lying on the line between $(0, 0)$ and $(1, 1)$. The componentwise mean of the data gives $(0.375, 0.625)$ marked

with $\times$ in Fig. 1a. Then the componentwise median gives $(0.25, 0.75)$ as marked with $+$ in Fig. 1a. Finally, the measure defined in (3.4) gives the geometric median $\overline{Y} = (0.4996, 0.5004)$ marked with $\square$ in Fig. 1a, where the smooth landscape for the merit function M is shown in Fig. 1b. Because of the natural result obtained by the geometric median in Fig. 1a, (3.4) will be used here for the ℓ_1 vector mean.

To compute the ℓ_1 mean of (3.4) the following primal-dual iteration is used. For $\tau > 0$ compute iteratively $D_l \in \mathbb{R}^{m\times n}$ and $\boldsymbol{\mu}_l \in \mathbb{R}^m$ by

$$D_{l+1}\hat{\boldsymbol{e}}_j = \frac{D_l\hat{\boldsymbol{e}}_j + \tau(\boldsymbol{\mu}_l - Y\hat{\boldsymbol{e}}_j)}{1+\tau\|\boldsymbol{\mu}_l - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}}, \quad j = 1,\ldots,n \tag{3.6}$$

$$\boldsymbol{\mu}_{l+1} = \sum_{j=1}^{n} \frac{(D_l - \tau Y)\hat{\boldsymbol{e}}_j}{1+\tau\|\boldsymbol{\mu}_l - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}} \Big/ \sum_{j=1}^{n} \frac{-\tau}{1+\tau\|\boldsymbol{\mu}_l - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}}. \tag{3.7}$$

The motivation for this iteration and its convergence analysis are given in Section 6. The ℓ_1-mean is given by taking the limit,

$$\overline{Y} = \lim_{l\to\infty} \boldsymbol{\mu}_l. \tag{3.8}$$

After these calculations have been completed, the centered data are given be $Y_{\rm c} = Y - \overline{Y}$, the counterpart to (2.6) with (3.4) replacing (2.7).

4 ℓ_1 Approach to PCA

To present our approach we start with some preliminaries. Unless otherwise specified, it is assumed that $m \le n$. With V and Λ in (2.10) given in terms of components as $V = \{\hat{\boldsymbol{v}}_i\}_{i=1}^m$ and $\Lambda = \text{diag}\{\lambda_i\}_{i=1}^m$, respectively, define

$$V_k = \{\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_k\} \in \mathbb{R}^{m\times k} \tag{4.1}$$

and the projected data

$$Y_k = (I - V_{k-1}V_{k-1}^{\rm T})Y_{\rm c}, \quad k = 2,\ldots,m, \quad Y_1 = Y_{\rm c}. \tag{4.2}$$

Let $S_k = \mathcal{R}(Y_k)$ where $\mathcal{R}$ denotes the range. For convenience, it is assumed here that the data $Y_{\rm c}$ have maximal rank so that $S_1 = \mathbb{R}^m$. Then $\boldsymbol{0} = \boldsymbol{v}^{\rm T}(I - V_{k-1}V_{k-1}^{\rm T})Y_{\rm c} = \boldsymbol{v}^{\rm T}Y_k$ is equivalent to $\boldsymbol{v} = V_{k-1}V_{k-1}^{\rm T}\boldsymbol{v}$, which is equivalent to $\boldsymbol{v} = \boldsymbol{0}$ exactly when $V_{k-1}^{\rm T}\boldsymbol{v} = \boldsymbol{0}$. Hence,

$$S_k = \mathcal{R}(V_{k-1})^{\perp}, \quad k = 2,\ldots,m, \quad S_1 = \mathbb{R}^m. \tag{4.3}$$

Before presenting the proposed robust measure for determining visually natural data axes, a motivation is given by reformulating the ℓ_2 eigenspace decomposition in (2.10) in terms of a least squares fit of an axis system to the cloud of data points. Given S_k, let the k-th column of V and the k-th diagonal entry of Λ be determined inductively by the regression

$$\hat{\boldsymbol{v}}_k = \frac{\boldsymbol{v}_k}{\|\boldsymbol{v}_k\|_{\ell_2}}, \quad \boldsymbol{v}_k = \underset{\boldsymbol{v}\in S_k}{\text{argmin}}\, \tilde{H}_k(\boldsymbol{v}), \quad \lambda_k = \|\hat{\boldsymbol{v}}_k^{\rm T}Y_k/\sqrt{n}\|_{\ell_2}^2, \quad k = 1,\ldots,m \tag{4.4}$$

where

$$\tilde{H}_k(\boldsymbol{v}) = \sum_{j=1}^{n} \left\| \left(\frac{\boldsymbol{v}\boldsymbol{v}^{\rm T}}{\|\boldsymbol{v}\|_{\ell_2}^2} - I \right) Y_k\hat{\boldsymbol{e}}_j \right\|_{\ell_2}^2, \quad \boldsymbol{v} \ne \boldsymbol{0}, \quad \tilde{H}_k(\boldsymbol{0}) = \|Y_k\|_{\rm F}^2. \tag{4.5}$$

Since minimizing

$$\begin{aligned} \tilde{H}_k(\hat{\boldsymbol{v}}) &= \sum_{j=1}^{n} \hat{\boldsymbol{e}}_j^{\rm T}Y_k^{\rm T}(\hat{\boldsymbol{v}}\hat{\boldsymbol{v}}^{\rm T} - I)^{\rm T}(\hat{\boldsymbol{v}}\hat{\boldsymbol{v}}^{\rm T} - I)Y_k\hat{\boldsymbol{e}}_j = \sum_{j=1}^{n} \hat{\boldsymbol{e}}_j^{\rm T}Y_k^{\rm T}(I - \hat{\boldsymbol{v}}\hat{\boldsymbol{v}}^{\rm T})Y_k\hat{\boldsymbol{e}}_j \\ &= \sum_{j=1}^{n} \left[|Y_k\hat{\boldsymbol{e}}_j|^2 - |\hat{\boldsymbol{e}}_j^{\rm T}Y_k^{\rm T}\hat{\boldsymbol{v}}|^2 \right] = \|Y_k\|_{\rm F}^2 - \|Y_k^{\rm T}\hat{\boldsymbol{v}}\|_{\ell_2}^2 \end{aligned} \tag{4.6}$$

over $\hat{\boldsymbol{v}} \in S_k$ with $\|\hat{\boldsymbol{v}}\|_{\ell_2} = 1$ is equivalent to maximizing the Rayleigh quotient $\boldsymbol{v}^{\mathrm{T}} Y_k Y_k^{\mathrm{T}} \boldsymbol{v} / \boldsymbol{v}^{\mathrm{T}} \boldsymbol{v} = \|Y_k^{\mathrm{T}} \boldsymbol{v}\|_{\ell_2}^2 / \|\boldsymbol{v}\|_{\ell_2}^2$ over the same set, $\hat{\boldsymbol{v}}_k$ in (4.4) is the eigenvector of $\frac{1}{n} Y_{\mathrm{c}} Y_{\mathrm{c}}^{\mathrm{T}}$ with the k-th largest eigenvalue $\lambda_k = \|\hat{\boldsymbol{v}}_k^{\mathrm{T}} Y_{\mathrm{c}} / \sqrt{n}\|_{\ell_2}^2 = \|\hat{\boldsymbol{v}}_k^{\mathrm{T}} Y_k / \sqrt{n}\|_{\ell_2}^2$ as shown in (4.4).

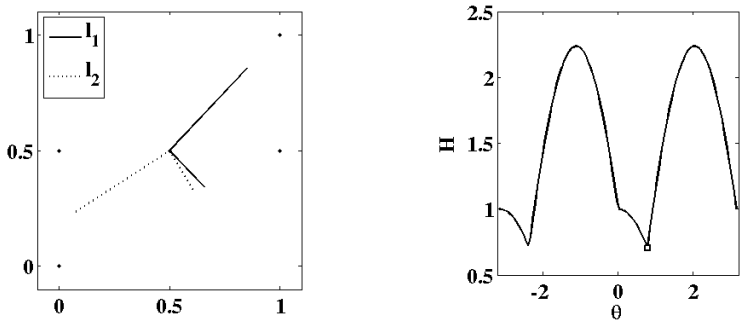
We now aim for an appropriate ℓ_1-variant of (4.4) - (4.5). The proposed approach to determine the orthogonal matrix V and the diagonal matrix Λ in (2.5) is to replace the sum of squared norms in (4.5) with a sum of norms in (4.8) below. The approach is remniscent of (3.4) in the sense that while a geometric median *point* is selected by (3.4), a geometric median *line* is determined by (4.7). As for (4.5), let V_k be given by (4.1) and Y_k by (4.2). Then given S_k according to (4.3), let $\boldsymbol{v}_k$ and λ_k be determined inductively by

$$\hat{\boldsymbol{v}}_k = \frac{\boldsymbol{v}_k}{\|\boldsymbol{v}_k\|_{\ell_2}}, \quad \boldsymbol{v}_k = \operatorname*{argmin}_{\boldsymbol{v} \in S_k} H_k(\boldsymbol{v}), \quad \lambda_k = \|\hat{\boldsymbol{v}}_k^{\mathrm{T}} Y_k / \sqrt{n}\|_{\ell_1}, \quad k = 1, \ldots, m \tag{4.7}$$

where

$$H_k(\boldsymbol{v}) = \sum_{j=1}^{n} \left\| \left(\frac{\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}}{\|\boldsymbol{v}\|_{\ell_2}^2} - I \right) Y_k \hat{\boldsymbol{e}}_j \right\|_{\ell_2}, \quad \boldsymbol{v} \neq \boldsymbol{0}, \quad H_k(\boldsymbol{0}) = \sum_{j=1}^{n} \|Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}. \tag{4.8}$$

The robustness of the ℓ_1 measure in H_k in relation to the ℓ_2 measure in $\tilde{H}_k$ is highlighted by the following simple example, which is illustrated in Fig. 2. Here the data are given by



(a) data and ℓ_1 and ℓ_2 data axes (b) landscape of the merit function $H_1(\theta)$

Figure 2: The ℓ_1 (solid) and ℓ_2 (dotted) data axes are compared in (a), where the data of (4.9) are shown with $\cdot$. Shown in (b) is the landscape of the merit function $H_1(\theta) = H_1(\hat{\boldsymbol{v}}(\theta))$ in (4.8) with $\hat{\boldsymbol{v}}(\theta) = (\cos(\theta), \sin(\theta)$, which is minimized at $\square$.

$$Y = \begin{bmatrix} 0 & \frac{1}{2} & 1 & 0 & 1 \\ 0 & \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \tag{4.9}$$

marked with $\cdot$ in Fig. 2a. The points $(0, \frac{1}{2}), (1, \frac{1}{2})$ may be regarded as outliers from points otherwise lying on the line between $(0, 0)$ and $(1, 1)$. The ℓ_2 data axes are given by (4.4) and are shown in Fig. 2a as dotted line segments. The ℓ_1 data axes are given by (4.7) and are shown in Fig. 2a as solid line segments. The landscape for the merit function $H_1(\theta) = H_1(\hat{\boldsymbol{v}}(\theta))$ of (4.8) with $\hat{\boldsymbol{v}}(\theta) = (\cos(\theta), \sin(\theta)$ is shown in Fig. 2b, and the minimum is marked with $\square$; see Remark 2 concerning the regularity of the merit function.

The data axes defined by (4.7) are computed by the following scheme. For $\tau > 0$ and $\rho > 0$ compute iteratively $D_l \hat{\boldsymbol{e}}_j \in S_k$, $j = 1, \ldots, n$, and $\hat{\mathbf{v}}_l \in S_k$ with $\|D_0 \hat{\boldsymbol{e}}_j\|_{\ell_2} \leq 1, j = 1, \ldots, n$, and $\|\hat{\mathbf{v}}_0\|_{\ell_2} = 1$,

$$D_{l+1} \hat{\boldsymbol{e}}_j = \frac{(\hat{\mathbf{v}}_l \hat{\mathbf{v}}_l^{\mathrm{T}} - I)(\tau Y_k \hat{\boldsymbol{e}}_j - D_l \hat{\boldsymbol{e}}_j)}{1 + \tau \|(\hat{\mathbf{v}}_l \hat{\mathbf{v}}_l^{\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}}, \quad j = 1, \ldots, n \tag{4.10}$$

$$\hat{\mathbf{v}}_{l+1} = \frac{\mathbf{v}_{l+1}}{\|\mathbf{v}_{l+1}\|_{\ell_2}} \quad \text{with} \quad \mathbf{v}_{l+1} = \hat{\mathbf{v}}_l - \rho \sum_{j=1}^{n} \frac{(\hat{\mathbf{v}}_l^{\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j)(\tau Y_k \hat{\boldsymbol{e}}_j - D_l \hat{\boldsymbol{e}}_j)}{1 + \tau \|(\hat{\mathbf{v}}_l \hat{\mathbf{v}}_l^{\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}}. \tag{4.11}$$

The motivation for this iteration and its convergence analysis are given in Section 7. Then the k-th components of V and Λ are given respectively by taking the limit,

$$\boldsymbol{v}_k = \lim_{l\to\infty} \hat{\mathbf{v}}_l \tag{4.12}$$

and setting

$$\lambda_k = R_k(\boldsymbol{v}_k). \tag{4.13}$$

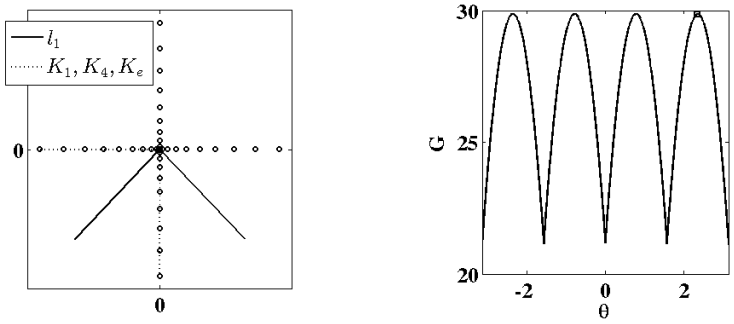
After these calculations have been completed for $k = 1, \ldots, m$, the ℓ_1 sphered data are given by $Y_{\rm s} = \Lambda^{-\frac{1}{2}} V^{\rm T} Y_{\rm c}$, the counterpart to (2.8).

Since the minimization problems (4.7) are performed over progressively smaller subspaces S_k, it follows that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. Thus, as in Section 2, the ℓ_1-variation λ_i of the data $Y_{\rm c}$ along the axis $\boldsymbol{v}_i$ is larger than the ℓ_1-variation λ_j along the axis $\boldsymbol{v}_j$ for $i < j$. To select only the $r < m$ components with respect to which the data have the most ℓ_1-variation, define the projected data $Y_P \approx Y$ by the counterpart to (2.11) where the matrices $\overline{Y}$, V and Λ are now determined by ℓ_1 measures while the projector $P \in \mathbb{R}^{r\times m}$ is defined as before with entries $P_{i,j} = \delta_{i,j}$.

In case the data are undersampled and $n < m$, the steps outlined in this section must be modified as follows. Let $\overline{Y}$ and $Y_{\rm c}$ be determined as described following (3.4). Then $\hat{V} \in \mathbb{R}^{m\times n}$ is determined by solving (4.7) but for $k = 1, \ldots, n$ where S_1 is the range of $Y_{\rm c}$. With $\lambda_k = R_k(\boldsymbol{v}_k)$, set $\hat{\Lambda} = \text{diag}\{\lambda_k\}_{k=1}^n$. The sphered data are then given by $\hat{Y}_{\rm s} = \hat{\Lambda}^{-\frac{1}{2}} \hat{V}^{\rm T} Y_{\rm c} \in \mathbb{R}^{n\times n}$. Finally, for a projector $P \in \mathbb{R}^{n\times n}$, (2.11) becomes $Y_P = \overline{Y} + \hat{V}\hat{\Lambda}^{\frac{1}{2}} P^{\rm T} P \hat{\Lambda}^{-\frac{1}{2}} \hat{V}^{\rm T}(Y - \overline{Y})$.

5 ℓ_1 Approach to ICA

While the Kurtosis has been used as a measure of Gaussianity in (2.18) to determine independent components, an alternative measure of independence is proposed here which is more robust in the presence of outliers. This approach targets independence directly in a manner which can be illustrated in terms of the example shown below in Fig. 3. Here the data are given by



(a) data and data axes by ℓ_1, K_1, K_4 and $K_{\rm e}$ (b) landscape of merit function $\|U(\theta)Y\|_{\ell_1}$

Figure 3: The ℓ_1 (solid) and K_2, K_4 and $K_{\rm e}$ (dotted, i.e., identical) data axes where obtained by maximizing the measures in (5.5). The results are compared in (a), where the data of (5.10) are shown with $\cdot$. Shown in (b) is the landscape of the merit function $\|U(\theta)Y\|_{\ell_1}$ which is maximized at $\square$.

$$Y = \begin{bmatrix} -1 & \ldots & \sigma(\frac{i}{m})|\frac{i}{m}|^k & \ldots & 1 & 0 & \ldots & & \ldots & 0 \\ 0 & \ldots & & \ldots & 0 & -1 & \ldots & \sigma(\frac{i}{m})|\frac{i}{m}|^k & \ldots & 1 \end{bmatrix}, \quad \begin{matrix} m = 20, k = 2 \\ i = -m, \ldots, m \end{matrix} \tag{5.1}$$

where $\sigma(t) = \text{sign}(t)$. Let these data represent a realization of a random vector $\mathrm{y} \in \mathbb{R}^2$ satisfying

$$\mathrm{P}(\mathrm{y} = Y\hat{\boldsymbol{e}}_i) = \mathrm{P}(\mathrm{y} = Y\hat{\boldsymbol{e}}_j) = \frac{1}{2m}, \quad (\hat{\boldsymbol{e}}_i)_j = \delta_{ij} \tag{5.2}$$

where P denotes the probability. In order that the rotation dependent random vector

$$\mathrm{x}(\theta) = \{\mathrm{x}_i(\theta)\}_{i=1}^2, \quad \mathrm{x}(\theta) = U(\theta)\mathrm{y}, \quad U(\theta) = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \tag{5.3}$$

satisfy the independence condition,

$$\mathrm{P}(\mathrm{x}_1(\theta) = \alpha \text{ and } \mathrm{x}_2(\theta) = \beta) = \mathrm{P}(\mathrm{x}_1(\theta) = \alpha) \cdot \mathrm{P}(\mathrm{x}_2(\theta) = \beta), \quad \forall \alpha, \beta \in \mathbb{R} \tag{5.4}$$

a rotation angle $\theta = \pi/4 + k\pi/2$, $k \in \mathbb{Z}$, must be chosen. For the determination of the proper rotation, four different measures of independence are compared in Fig. 3,

$$\|U(\theta)Y\|_{\ell_1}, \quad K_1^2(U(\theta)Y), \quad K_4^2(U(\theta)Y), \quad \text{and} \quad K_{\mathrm{e}}^2(U(\theta)Y), \tag{5.5}$$

where K_m is given by (2.13) and (see [10])

$$K_{\mathrm{e}}(\mathrm{x}) = \mathrm{E}[\exp(-|\mathrm{x} - \mathrm{E}[\mathrm{x}]|^2)/2] - 1/\sqrt{1 + \mathrm{E}[|\mathrm{x} - \mathrm{E}[\mathrm{x}]|^2]} \tag{5.6}$$

The data axes obtained by maximizing the last three measures in (5.5) are identical and are shown in Fig. 7a as dotted line segments. The data axes obtained by maximizing the first measure in (5.5) are shown in Fig. 7a as solid line segments. The landscape for the merit function $\|U(\theta)Y\|_{\ell_1}$ of (5.5) is shown in Fig. 7b and the maximum is marked with □. Only the first measure in (5.5) is maximized at a desired angle as shown in the landscape of Fig. 7b. All other measures are maximized at a multiple of π. On the basis of this example, the measure shown below in (5.8) is proposed to determine the rotation matrix U of (2.14).

To achieve maximally independent rows of $X_{\mathrm{c}} = UY_{\mathrm{s}}$, the rows of the orthogonal matrix $U = \{\hat{\boldsymbol{u}}_i^{\mathrm{T}}\}_{i=1}^m$ are determined as follows. Define $U_l = \{\hat{\boldsymbol{u}}_1, \ldots, \hat{\boldsymbol{u}}_l\}^{\mathrm{T}}$ as in (2.15) and the projected data $Y_l = (I - U_{l-1}^{\mathrm{T}}U_{l-1})Y_{\mathrm{s}}$, $l = 2, \ldots, m$, $Y_1 = Y_{\mathrm{s}}$, as in (2.16). Let $T_l = \mathcal{R}(Y_l)$ where $\mathcal{R}$ denotes the range. For convenience, it is assumed here that the data Y_{s} have maximal rank so that $T_1 = \mathbb{R}^m$. Then $\mathbf{0} = \boldsymbol{u}^{\mathrm{T}}(I - U_{l-1}^{\mathrm{T}}U_{l-1})Y_{\mathrm{s}} = \boldsymbol{u}^{\mathrm{T}}Y_l$ is equivalent to $\boldsymbol{u} = U_{l-1}^{\mathrm{T}}U_{l-1}\boldsymbol{u}$, which is equivalent to $\boldsymbol{u} = \mathbf{0}$ exactly when $U_{l-1}\boldsymbol{u} = \mathbf{0}$. Hence,

$$T_l = \mathcal{R}(U_{l-1}^{\mathrm{T}})^{\perp}, \quad l = 2, \ldots, m, \quad T_1 = \mathbb{R}^m. \tag{5.7}$$

Given T_l, let the l-th column of U^{T} be determined inductively by

$$\hat{\boldsymbol{u}}_l = \frac{\mathbf{u}_l}{\|\mathbf{u}_l\|_{\ell_2}}, \quad \mathbf{u}_l = \operatorname*{argmax}_{\mathbf{u} \in T_l} G_l(\mathbf{u}) \tag{5.8}$$

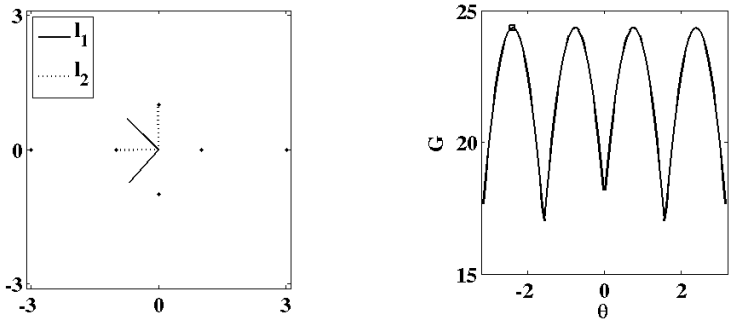
where

$$G_l(\mathbf{u}) = \frac{\|\mathbf{u}^{\mathrm{T}}Y_l\|_{\ell_1}}{\|\mathbf{u}\|_{\ell_2}}, \quad \mathbf{u} \neq \mathbf{0}, \quad G_l(\mathbf{0}) = 0. \tag{5.9}$$

In this way the rows of U are determined sequentially so that the earlier components of x are more strongly separated from other later components. Alternatively, all components of x may be estimated with roughly equal quality by determining all rows of U simultaneously through maximizing a sum of functionals of the form (5.9) for each row under the constraint that U be orthogonal. Once matrices U, Λ and V are determined, the source samples are estimated according to (2.4) and (2.5).

The robustness of the measure G_l in (5.9) in relation to the measure F in (2.20) is highlighted by the following simple example, which is illustrated in Fig. 4. Here the data Y are given by

$$\begin{aligned} Y_x &= \begin{bmatrix} +1 & +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\ Y_y &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} Y_x, \quad Y_{\mathrm{o}} = \begin{bmatrix} 3 & -3 \\ 0 & 0 \end{bmatrix}, \quad Y = \begin{bmatrix} Y_x & Y_y & Y_{\mathrm{o}} \end{bmatrix} \end{aligned} \tag{5.10}$$



(a) data and data axes by ℓ_1 and K_1 (b) landscape of merit function $\|U(\theta)Y\|_{\ell_1}$

Figure 4: The ℓ_1 (solid) and ℓ_2 (dotted) data axes are compared in (a), where the data of (5.10) are shown with $\cdot$. Shown in (b) is the landscape of the merit function G_1 in (5.9).

marked with $\cdot$ in Fig. 4a, and the points $(\pm 3, 0)$ may be regarded as outliers from points otherwise lying at the diamond vertices $\{(0, \pm 1), (\pm 1, 0)\}$. The ℓ_2 data axes are given by (2.18) and are shown in Fig. 4a as dotted line segments. The ℓ_1 data axes are given by (5.8) and are shown in Fig. 4a as solid line segments, where the landscape for the merit function G_1 of (5.9) is shown in Fig. 4b and the maximum is marked with $\square$.

The vectors defined by (5.8) are computed by the following scheme. For $\tau > 0$ compute iteratively $\hat{\mathbf{u}}_k \in T_l$ with $\|\hat{\mathbf{u}}_0\|_{\ell_2} = 1$,

$$\hat{\mathbf{u}}_{k+1} = \frac{\mathbf{u}_{k+1}}{\|\mathbf{u}_{k+1}\|_{\ell_2}} \quad \text{with} \quad \mathbf{u}_{k+1} = \hat{\mathbf{u}}_k + \tau \left[Y_l \, \sigma(Y_l^{\mathrm{T}} \hat{\mathbf{u}}_k) - \hat{\mathbf{u}}_k \|\hat{\mathbf{u}}_k^{\mathrm{T}} Y_l\|_{\ell_1} \right] \tag{5.11}$$

where

$$\sigma(t) = \operatorname{sign}(t) \quad \text{for} \quad t \in \mathbb{R}, \qquad \sigma(\boldsymbol{v}) = \{\sigma(v_j)\}_{j=1}^n \quad \text{for} \quad \boldsymbol{v} = \{v_j\}_{j=1}^n \in \mathbb{R}^n. \tag{5.12}$$

The motivation for this iteration and its convergence analysis are given in Section 8. The l-th column of U^{T} is given by taking the limit,

$$\hat{\boldsymbol{u}}_l = \lim_{k \to \infty} \hat{\mathbf{u}}_k \tag{5.13}$$

After these calculations have been completed for $l = 1, \ldots, m$, the ℓ_1 maximally independent data are given by $X_{\mathrm{c}} = UY_{\mathrm{s}}$, the counterpart to (2.14).

To select the $r < m$ desired independent components $\{q_1, \ldots, q_r\} \subset \{1, \ldots, m\}$, define the projected data $Y_Q \approx Y$ by the counterpart to (2.21) where the matrices $\overline{Y}$, V, Λ and U are now determined by ℓ_1 measures while the projector $Q \in \mathbb{R}^{r \times m}$ is defined as before with entries $Q_{ij} = \delta_{q_i, j}$.

In case the data are undersampled and $n < m$, let $\hat{V} \in \mathbb{R}^{m \times n}$, $\hat{Y}_{\mathrm{s}}, \hat{\Lambda} \in \mathbb{R}^{n \times n}$ be given as described at the end of Section 3. Then, the sphered data $\hat{Y}_{\mathrm{s}}$ are transformed by the rotation matrix $\hat{U} \in \mathbb{R}^{n \times n}$ maximizing independence of the rows of $\hat{X}_{\mathrm{c}} = \hat{U}\hat{Y}_{\mathrm{s}} \in \mathbb{R}^{n \times n}$. Thus, $\hat{X} = \hat{U}\hat{\Lambda}^{-\frac{1}{2}}\hat{V}^{\mathrm{T}}Y$ gives the maximum number of independent components which can be determined from the undersampled data. Finally, for a projector $Q \in \mathbb{R}^{n \times n}$, (2.21) becomes $Y_Q = \overline{Y} + \hat{V}\hat{\Lambda}^{\frac{1}{2}}\hat{U}^{\mathrm{T}}Q^{\mathrm{T}}Q\hat{U}\hat{\Lambda}^{-\frac{1}{2}}\hat{V}^{\mathrm{T}}(Y - \overline{Y})$.

6 Convergence of the Iterative Scheme for the ℓ_1 Mean

The analysis of the scheme (3.6) – (3.7) begins by establishing basic properties for the minimization problem (3.4), i.e., the determination of the geometric median. As indicated in Section 1, a proof of uniqueness of the geometric median is provided here which is shorter than that found in [15] or [13] and is based on strict convexity of the functional M. Also, a possibly novel characterization of a solution is provided in case the columns of Y are colinear, which means that Y can be expressed in the form

$$Y = \boldsymbol{a}\boldsymbol{e}^{\mathrm{T}} + \boldsymbol{b}\boldsymbol{y}^{\mathrm{T}}, \text{where } \boldsymbol{e} = (1, \ldots, 1)^{\mathrm{T}}, \ \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^m, \ \boldsymbol{e}, \boldsymbol{y} \in \mathbb{R}^n. \tag{6.1}$$

Lemma 1 *If the columns of $Y \in \mathbb{R}^{m\times n}$ are not colinear, then M is strictly convex.*

Proof: The mapping M is the sum of convex mappings and hence convex itself. If M is not strictly convex then there are vectors $\boldsymbol{\mu_1}, \boldsymbol{\mu_2} \in \mathbb{R}^m$ and $\overline{\boldsymbol{\mu}} = \overline{\alpha}\boldsymbol{\mu_1} + (1-\overline{\alpha})\boldsymbol{\mu_2},$ with $\overline{\alpha} \in (0,1)$ such that

$$M(\overline{\boldsymbol{\mu}}) = \overline{\alpha}M(\boldsymbol{\mu_1}) + (1-\overline{\alpha})M(\boldsymbol{\mu_2}).$$

We claim that in this case the convex function $\alpha \to M(\alpha\boldsymbol{\mu_1} + (1-\alpha)\boldsymbol{\mu_2})$, $\alpha \in [0,1]$ is the affine mapping $h(\alpha) = \alpha M(\boldsymbol{\mu_1}) + (1-\alpha)M(\boldsymbol{\mu}_2)$. If this is not the case, then, using convexity, there exists $\tilde{\alpha} \in (0,1)$ such that for $\tilde{\boldsymbol{\mu}} = \tilde{\alpha}\boldsymbol{\mu_1} + (1-\tilde{\alpha})\boldsymbol{\mu_2}$ we have $M(\tilde{\boldsymbol{\mu}}) < h(\tilde{\alpha})$. Assume at first that $\tilde{\alpha} < \overline{\alpha}$. Then $\overline{\boldsymbol{\mu}} = \beta\tilde{\boldsymbol{\mu}} + (1-\beta)\boldsymbol{\mu_2}$ for some $\beta \in (0,1)$ and we have $M(\overline{\boldsymbol{\mu}}) \le \beta M(\tilde{\boldsymbol{\mu}}) + (1-\beta)M(\boldsymbol{\mu_2}) < \beta h(\tilde{\alpha}) + (1-\beta)M(\boldsymbol{\mu_2}) = M(\overline{\boldsymbol{\mu}})$ which is a contradiction. The case $\tilde{\alpha} > \overline{\alpha}$ can be treated analogously, and thus $\alpha \to M(\alpha\boldsymbol{\mu_1} + (1-\alpha)\boldsymbol{\mu_2})$, $\alpha \in (0,1)$, is affine.

Let $\hat{\alpha} \in (0,1)$ be such that $\hat{\boldsymbol{\mu}} = \hat{\alpha}\boldsymbol{\mu_1} + (1-\hat{\alpha})\boldsymbol{\mu_2}$ does not coincide with any of the columns $Y\hat{\boldsymbol{e}}_i$. Then we find the Hessian

$$\nabla^2 M(\hat{\boldsymbol{\mu}}) = \sum_{i=1}^{n} \frac{1}{\|\hat{\boldsymbol{\mu}} - Y\hat{\boldsymbol{e}}_i\|_{\ell_2}^3} \big(\|\hat{\boldsymbol{\mu}} - Y\hat{\boldsymbol{e}}_i)\|_{\ell_2}^2 I - (\hat{\boldsymbol{\mu}} - Y\hat{\boldsymbol{e}}_i)(\hat{\boldsymbol{\mu}} - Y\hat{\boldsymbol{e}}_i)^{\mathrm{T}}\big), \tag{6.2}$$

and hence for any $\boldsymbol{x} \in \mathbb{R}^m$

$$\boldsymbol{x}^{\mathrm{T}}\nabla^2 M(\hat{\boldsymbol{\mu}})\boldsymbol{x} = \sum_{i=1}^{n} \frac{1}{\|\hat{\boldsymbol{\mu}} - Y\hat{\boldsymbol{e}}_i\|_{\ell_2}^3} \big(\|\hat{\boldsymbol{\mu}} - Y\hat{\boldsymbol{e}}_i)\|_{\ell_2}^2 \|\boldsymbol{x}\|_{\ell_2}^2 - (\hat{\boldsymbol{\mu}} - Y\hat{\boldsymbol{e}}_i)^{\mathrm{T}}\boldsymbol{x})^2\big). \tag{6.3}$$

Note that $|(\hat{\boldsymbol{\mu}} - Y\hat{\boldsymbol{e}}_i)^{\mathrm{T}}\boldsymbol{x}| \le \|\hat{\boldsymbol{\mu}} - Y\hat{\boldsymbol{e}}_i\|_{\ell_2} \|\boldsymbol{x}\|_{\ell_2}$. If $|(\hat{\boldsymbol{\mu}} - Y\hat{\boldsymbol{e}}_i)^{\mathrm{T}}\boldsymbol{x}| = \|\hat{\boldsymbol{\mu}} - Y\hat{\boldsymbol{e}}_i\| \|\boldsymbol{x}\|$ for all $i = 1, \ldots, n$, then there exist $b_i \in \mathbb{R}$ such that $\hat{\boldsymbol{\mu}} - Y\hat{\boldsymbol{e}}_i = b_i\boldsymbol{x}$, for $i = 1, \ldots, n$. Thus there exists $\boldsymbol{b} \in \mathbb{R}^n$ such that $Y = \boldsymbol{\mu}\boldsymbol{e}^{\mathrm{T}} - \boldsymbol{b}\boldsymbol{x}^{\mathrm{T}}$ which contradicts the assumption. Hence there exists at least one index i such that $|(\hat{\boldsymbol{\mu}} - Y\hat{\boldsymbol{e}}_i)^{\mathrm{T}}\boldsymbol{x}| < \|\hat{\boldsymbol{\mu}} - Y\hat{\boldsymbol{e}}_i\|_{\ell_2} \|\boldsymbol{x}\|_{\ell_2}$ and thus $\boldsymbol{x}^{\mathrm{T}}\nabla^2 M(\hat{\boldsymbol{\mu}})\boldsymbol{x} > 0$. This contradicts that $\alpha \to M(\alpha\boldsymbol{\mu_1} + (1-\alpha)\boldsymbol{\mu_2})$ is affine at $\hat{\alpha}$. ∎

Lemma 2 *If the columns of $Y \in \mathbb{R}^{m\times n}$ are colinear, then M is minimized (not necessarily uniquely) by $\boldsymbol{\mu}^\star = \boldsymbol{a} + \boldsymbol{b} \cdot$ median $(\boldsymbol{y})$. If the columns of Y are not colinear, there exists a unique $\boldsymbol{\mu}^\star \in \mathbb{R}^m$ minimizing M.*

Proof: Existence of a solution $\boldsymbol{\mu}^* \in \mathbb{R}^m$ follows by standard subsequential limit arguments. If the columns of Y are not colinear, uniqueness of the solution $\boldsymbol{\mu}^\star$ follows from strict convexity of $\boldsymbol{\mu} \to M(\boldsymbol{\mu})$.

Suppose next that the columns of Y are colinear so that $Y = \boldsymbol{a}\boldsymbol{e}^{\mathrm{T}} + \boldsymbol{b}\boldsymbol{y}^{\mathrm{T}}$. Set $\nu = \|\boldsymbol{b}\|_{\ell_2}$ and $\boldsymbol{w} = \boldsymbol{b} + \nu\hat{\boldsymbol{e}}_1$ where $(\hat{\boldsymbol{e}}_1)_i = \delta_{i1}$. Then the Householder transformation

$$U = I - 2\frac{\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}}{\|\boldsymbol{w}\|_{\ell_2}^2} \tag{6.4}$$

is orthogonal and satisfies

$$U\boldsymbol{b} = -\nu\hat{\boldsymbol{e}}_1 \tag{6.5}$$

as well as $\|U\boldsymbol{x}\|_{\ell_2} = \|\boldsymbol{x}\|_{\ell_2}$, $\forall \boldsymbol{x} \in \mathbb{R}^m$. Let an arbitrary $\boldsymbol{\mu} \in \mathbb{R}^m$ be represented as

$$\boldsymbol{\mu} = \boldsymbol{a} + x\boldsymbol{b} + \tilde{\boldsymbol{b}}, \quad x = (\boldsymbol{\mu} - \boldsymbol{a})^{\mathrm{T}}\boldsymbol{b}/\|\boldsymbol{b}\|_{\ell_2}, \quad \boldsymbol{b}^{\mathrm{T}}\tilde{\boldsymbol{b}} = 0 \tag{6.6}$$

and note that

$$(U\boldsymbol{b})^{\mathrm{T}}(U\tilde{\boldsymbol{b}}) = \boldsymbol{b}^{\mathrm{T}}U^{\mathrm{T}}U\tilde{\boldsymbol{b}} = \boldsymbol{b}^{\mathrm{T}}\tilde{\boldsymbol{b}} = 0. \tag{6.7}$$

With (6.5) – (6.7), the merit function can be written as

$$\begin{aligned} M(\boldsymbol{\mu}) &= \sum_{j=1}^{n} \|\boldsymbol{\mu} - \boldsymbol{a} - \boldsymbol{b}y_j\|_{\ell_2} = \sum_{j=1}^{n} \|U(\boldsymbol{\mu} - \boldsymbol{a} - \boldsymbol{b}y_j)\|_{\ell_2} = \sum_{j=1}^{n} \|U(\tilde{\boldsymbol{b}} + (x - y_j)\boldsymbol{b})\|_{\ell_2} \\ &= \sum_{j=1}^{n} \left[\|U\tilde{\boldsymbol{b}}\|_{\ell_2}^2 + 2(x - y_j)(U\boldsymbol{b})^{\mathrm{T}}(U\tilde{\boldsymbol{b}}) + |x - y_j|^2\nu^2\right]^{\frac{1}{2}} \\ &= \sum_{j=1}^{n} \left[\|U\tilde{\boldsymbol{b}}\|_{\ell_2}^2 + |x - y_j|^2\nu^2\right]^{\frac{1}{2}} \end{aligned} \tag{6.8}$$

It follows that for $\gamma = \text{median}(\boldsymbol{y})$,

$$M(\boldsymbol{\mu}) \geq \nu \sum_{j=1}^{n} |x - y_j| \geq \nu \sum_{j=1}^{n} |\gamma - y_j| = M(\boldsymbol{a} + \gamma\boldsymbol{b}) \tag{6.9}$$

and hence M is minimized at $\boldsymbol{a} + \gamma\boldsymbol{b}$. That the minimizer is not necessarily unique can be seen from the case that n is even and the components of $\boldsymbol{y}$ are distinct. Then let the components of $\boldsymbol{y}$ be sorted in ascending order, set $p = n/2$ and $x = ty_p + (1-t)y_{p+1}$ for $t \in [0,1]$, where $\gamma = (y_p + y_{p+1})/2$. It follows from

$$\begin{aligned} \sum_{j=1}^{n} |x - y_j| &= \sum_{j=1}^{p} ([ty_p + (1-t)y_{p+1}] - y_j) + \sum_{j=p+1}^{n} (y_j - [ty_p + (1-t)y_{p+1}]) \\ &= -\sum_{j=1}^{p} y_j + \sum_{j=p+1}^{n} y_j \end{aligned} \tag{6.10}$$

that M has the same value $M(\boldsymbol{a} + \gamma\boldsymbol{b})$ at all points $\boldsymbol{a} + \boldsymbol{b}[ty_p + (1-t)y_{p+1}]$, $t \in [0,1]$. ∎

Lemma 3 *The first-order necessary optimality condition for a minimizer $\boldsymbol{\mu}^\star$ of M over $\mathbb{R}^m$ is that there be $D^\star \in \mathbb{R}^{m\times n}$ satisfying,*

$$\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j = \|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2} D^\star\hat{\boldsymbol{e}}_j, \quad \|D^\star\hat{\boldsymbol{e}}_j\|_{\ell_2} \leq 1, \quad j = 1,\ldots,n, \quad \sum_{j=1}^{n} D^\star\hat{\boldsymbol{e}}_j = \boldsymbol{0}. \tag{6.11}$$

Proof: The necessary optimality condition for a minimizer $\boldsymbol{\mu}^\star$ is that $0 \in \partial M(\boldsymbol{\mu}^\star)$. By the chain-rule (see, e.g., [2], p. 233), the subdifferential of M is given by the sum of the respective subdifferentials,

$$\partial M(\boldsymbol{\mu}) = \sum_{j=1}^{n} \partial\|\boldsymbol{\mu} - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}. \tag{6.12}$$

Thus, there exist $\boldsymbol{d}_j^\star \in \partial\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}$, $j = 1,\ldots,n$, satisfying

$$\sum_{j=1}^{n} \boldsymbol{d}_j^\star = 0, \tag{6.13}$$

and

$$\partial\|\boldsymbol{\mu} - Y\hat{\boldsymbol{e}}_j\|_{\ell_2} = \begin{cases} \dfrac{\boldsymbol{\mu} - Y\hat{\boldsymbol{e}}_j}{\|\boldsymbol{\mu} - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}}, & \boldsymbol{\mu} \neq Y\hat{\boldsymbol{e}}_j \\ B(0,1), & \boldsymbol{\mu} = Y\hat{\boldsymbol{e}}_j, \end{cases} \tag{6.14}$$

where $B(0,1)$ is the unit ball. Combining these facts we have

$$(\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j) = \|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}\boldsymbol{d}_j^\star, \quad j = 1,\ldots,n \tag{6.15}$$

The claim (6.11) follows with $D^\star = \{\boldsymbol{d}_1^\star, \ldots, \boldsymbol{d}_n^\star\}$. ∎

Turning to the iteration (3.6) – (3.7) we observe that if convergence to a fixed point $\{D^\star, \boldsymbol{\mu}^\star\}$ can be guaranteed, then from (3.6) we have that

$$D^\star \hat{\boldsymbol{e}}_j \|\boldsymbol{\mu}^\star - Y_j\|_{\ell_2} = \boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j, \quad j = 1, \ldots, n. \tag{6.16}$$

From (3.7) we have

$$\sum_{j=1}^{n} \frac{D^\star \hat{\boldsymbol{e}}_j + \tau(\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j)}{1 + \tau\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}} = 0. \tag{6.17}$$

Combining (6.16) and (6.17) gives

$$0 = \sum_{j=1}^{n} \frac{D^\star \hat{\boldsymbol{e}}_j + \tau D^\star \hat{\boldsymbol{e}}_j \|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}}{1 + \tau\|\boldsymbol{\mu}^\star - Y_j\|_{\ell_2}} = \sum_{j=1}^{n} D^\star \hat{\boldsymbol{e}}_j. \tag{6.18}$$

Moreover, if $\|D_0\hat{\boldsymbol{e}}_j\|_{\ell_2} \leq 1, j = 1, \ldots, n$, then the iterates $\{D_l\}$ also satisfy this bound, and hence $\|D^\star\hat{\boldsymbol{e}}_j\|_{\ell_2} \leq 1, j = 1, \ldots, n$. Thus $\{D^\star, \boldsymbol{\mu}^\star\}$ must satisfy the necessary optimality condition (6.11).

Theorem 1 *Suppose that the columns of $Y \in \mathbb{R}^{m\times n}$ are not colinear so that (6.1) does not hold. Let $\{D^\star, \boldsymbol{\mu}^\star\}$ satisfy (6.11) with $\boldsymbol{\mu}^\star \notin \{Y\hat{\boldsymbol{e}}_j\}_{j=1}^n$. Then $\{D^\star, \boldsymbol{\mu}^\star\}$ is a fixed point for the iteration (3.6) - (3.7) which is locally asymptotically stable for τ sufficiently large.*

Proof: Using (6.11) and substituting $D_l = D^\star$ on the right side of (3.6),

$$D_{l+1}\hat{\boldsymbol{e}}_j = \frac{D^\star \hat{\boldsymbol{e}}_j + \tau(\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j)}{1 + \tau\|\boldsymbol{\mu}_l - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}} = \frac{D^\star \hat{\boldsymbol{e}}_j + \tau\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2} D^\star\hat{\boldsymbol{e}}_j}{1 + \tau\|\boldsymbol{\mu}_l - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}} = D^\star\hat{\boldsymbol{e}}_j, \quad 1 \leq j \leq n \tag{6.19}$$

gives $D_{l+1} = D^\star$. Using this result together with (6.11) and setting $\boldsymbol{\mu}_l = \boldsymbol{\mu}^\star$ on the right side of (3.7),

$$\begin{aligned} \mathbf{0} &= \sum_{j=1}^{n} \frac{D^\star \hat{\boldsymbol{e}}_j + \tau(\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j) + +\tau(\boldsymbol{\mu}_{l+1} - \boldsymbol{\mu}^\star)}{1 + \tau\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}} \\ &= \sum_{j=1}^{n} D^\star\hat{\boldsymbol{e}}_j + \sum_{j=1}^{n} \frac{\tau(\boldsymbol{\mu}_{l+1} - \boldsymbol{\mu}^\star)}{1 + \tau\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}} = (\boldsymbol{\mu}_{l+1} - \boldsymbol{\mu}^\star) \sum_{j=1}^{n} \frac{\tau}{1 + \tau\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}} \end{aligned} \tag{6.20}$$

gives $\boldsymbol{\mu}_{l+1} = \boldsymbol{\mu}^\star$. Thus, $\{D^\star, \boldsymbol{\mu}^\star\}$ is a fixed point of the iteration (3.6) - (3.7). To establish the stability of the fixed point, define

$$\boldsymbol{F}_j(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_n, \boldsymbol{\mu}) = \frac{\boldsymbol{d}_j + \tau(\boldsymbol{\mu} - Y\hat{\boldsymbol{e}}_j)}{1 + \tau\|\boldsymbol{\mu} - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}}, \quad j = 1, \ldots, n \tag{6.21}$$

and

$$\boldsymbol{G}(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_n, \boldsymbol{\mu}) = \sum_{j=1}^{n} \frac{(\boldsymbol{d}_j - \tau Y\hat{\boldsymbol{e}}_j)}{1 + \tau\|\boldsymbol{\mu} - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}} \Big/ \sum_{j=1}^{n} \frac{-\tau}{1 + \tau\|\boldsymbol{\mu} - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}} \tag{6.22}$$

so that (3.6) – (3.7) is given by

$$D_{l+1}\hat{\boldsymbol{e}}_j = \boldsymbol{F}_j(D_l\hat{\boldsymbol{e}}_1, \ldots, D_l\hat{\boldsymbol{e}}_n, \boldsymbol{\mu}_l), \quad j = 1, \ldots, n, \quad \boldsymbol{\mu}_{l+1} = \boldsymbol{G}(D_{l+1}\hat{\boldsymbol{e}}_1, \ldots, D_{l+1}\hat{\boldsymbol{e}}_n, \boldsymbol{\mu}_l). \tag{6.23}$$

The claimed stability will follow once it is shown that the Jacobian of this mapping evaluated at $\{D^\star, \boldsymbol{\mu}^\star\} = \{\boldsymbol{d}_1^\star, \ldots, \boldsymbol{d}_n^\star, \boldsymbol{\mu}^\star\}$ has spectral radius less than 1 in when τ is sufficiently large. For (6.21),

$$\frac{\partial \boldsymbol{F}_j}{\partial \boldsymbol{d}}(\boldsymbol{d}_1^\star, \ldots, \boldsymbol{d}_n^\star, \boldsymbol{\mu}^\star) = \frac{I}{1 + \tau\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}} \overset{\tau\to\infty}{\longrightarrow} 0 \tag{6.24}$$

and

$$\begin{aligned}\frac{\partial \boldsymbol{F}_j}{\partial \boldsymbol{\mu}}(\boldsymbol{d}_1^\star,\ldots,\boldsymbol{d}_n^\star,\boldsymbol{\mu}^\star) &= \frac{\tau I}{1+\tau\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|} - \frac{\boldsymbol{d}_j^\star + \tau(\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j)}{[1+\tau\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}]^2}\frac{\tau(\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j)^{\mathrm{T}}}{\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}} \\ &\overset{\tau\to\infty}{\longrightarrow} \frac{1}{\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}}\left[I - \frac{(\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j)}{\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}}\frac{(\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j)^{\mathrm{T}}}{\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}}\right] =: A_j\end{aligned} \tag{6.25}$$

For (6.22),

$$\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{d}_j}(\boldsymbol{d}_1^\star,\ldots,\boldsymbol{d}_n^\star,\boldsymbol{\mu}^\star) = \sum_{i=1}^n \frac{I}{1+\tau\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i\|_{\ell_2}} \Big/ \sum_{j=1}^n \frac{-\tau}{1+\tau\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}} \overset{\tau\to\infty}{\longrightarrow} 0 \tag{6.26}$$

and

$$\begin{aligned}\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{\mu}}(\boldsymbol{d}_1^\star,\ldots,\boldsymbol{d}_n^\star,\boldsymbol{\mu}^\star) &= \sum_{i=1}^n \frac{(\boldsymbol{d}_i^\star - \tau Y\hat{\boldsymbol{e}}_i)}{[1+\tau\|\boldsymbol{\mu}^\star - \tau Y\hat{\boldsymbol{e}}_i\|_{\ell_2}]^2}\frac{\tau(\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i)^{\mathrm{T}}}{\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i\|_{\ell_2}} \Big/ \sum_{j=1}^n \frac{\tau}{1+\tau\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}} \\ &\quad + \underbrace{\sum_{i=1}^n \frac{(\boldsymbol{d}_i^\star - \tau Y\hat{\boldsymbol{e}}_i)}{1+\tau\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i\|_{\ell_2}}}_{=\sum_{i=1}^n \frac{-\tau\boldsymbol{\mu}^\star}{1+\tau\|\boldsymbol{\mu}^\star - Y\boldsymbol{e}_i\|_{\ell_2}}} \left[\sum_{j=1}^n \frac{-\tau}{[1+\tau\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}]^2}\frac{\tau(\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j)^{\mathrm{T}}}{\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}}\right] \Big/ \\ &\qquad \left[\sum_{k=1}^n \frac{-\tau}{1+\tau\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_k\|_{\ell_2}}\right]^2 \\ &= \sum_{i=1}^n \frac{[\boldsymbol{d}_i^\star + \tau(\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i)]}{[1+\tau\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i\|_{\ell_2}]^2}\frac{\tau(\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i)^{\mathrm{T}}}{\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i\|_{\ell_2}} \Big/ \sum_{j=1}^n \frac{\tau}{1+\tau\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}} \\ &\overset{\tau\to\infty}{\longrightarrow} \sum_{i=1}^n \frac{(\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i)(\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i)^{\mathrm{T}}}{\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i\|_{\ell_2}^3} \Big/ \sum_{j=1}^n \frac{1}{\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}} =: B\end{aligned} \tag{6.27}$$

Thus, the Jacobian satisfies

$$\frac{\partial(\boldsymbol{F}_1,\ldots,\boldsymbol{F}_n,\boldsymbol{G})}{\partial(\boldsymbol{d}_1,\ldots,\boldsymbol{d}_n,\boldsymbol{\mu})}(\boldsymbol{d}_1^\star,\ldots,\boldsymbol{d}_n^\star,\boldsymbol{\mu}^\star) \overset{\tau\to\infty}{\longrightarrow} \begin{bmatrix} 0 & \cdots & 0 & A_1 \\ \vdots & \ddots & & \vdots \\ 0 & & 0 & A_n \\ 0 & \cdots & 0 & B \end{bmatrix} =: J \tag{6.28}$$

The matrix $B \in \mathbb{R}^{m\times m}$ is clearly symmetric positive semi-definite, and it will now be shown that its spectrum lies in $[0,1)$. Suppose there is an $\boldsymbol{x} \in \mathbb{R}^m$ satisfying

$$\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i) = \|\boldsymbol{x}\|_{\ell_2}\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i\|_{\ell_2}, \quad i = 1,\ldots,n. \tag{6.29}$$

Then there exists $\boldsymbol{\alpha} = \{\alpha_i\}_{i=1}^n$ such that

$$\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i = \alpha_i \boldsymbol{x} \quad \Rightarrow \quad Y = \boldsymbol{\mu}^\star \boldsymbol{e}^{\mathrm{T}} - \boldsymbol{x}\boldsymbol{\alpha}^{\mathrm{T}} \tag{6.30}$$

violating the assumption that the columns of Y are not colinear. Thus, there can be no $\boldsymbol{x}$ satisfying (6.29). This result implies strict inequality in the following estimate:

$$\begin{aligned}\boldsymbol{x}^{\mathrm{T}}B\boldsymbol{x} &= \sum_{i=1}^n \frac{[\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i)]^2}{\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i\|_{\ell_2}^3} \Big/ \sum_{j=1}^n \frac{1}{\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}} \\ &< \sum_{i=1}^n \frac{\|\boldsymbol{x}\|_{\ell_2}^2\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i\|_{\ell_2}^2}{\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_i\|_{\ell_2}^3} \Big/ \sum_{j=1}^n \frac{1}{\|\boldsymbol{\mu}^\star - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}} = \|\boldsymbol{x}\|_{\ell_2}^2, \quad \forall \boldsymbol{x} \in \mathbb{R}^m\end{aligned} \tag{6.31}$$

Thus, the spectral radius of B is less than 1. Let a rotation matrix P be chosen so that $P^{\mathrm{T}}BP = \Lambda = \mathrm{diag}\{\lambda_i\}_{i=1}^m$ where $\lambda_i \in [0,1)$. Then the matrix J in (6.28) satisfies

$$\begin{bmatrix} I & \cdots & & 0 \\ \vdots & \ddots & & \vdots \\ 0 & & I & 0 \\ 0 & \cdots & 0 & P \end{bmatrix}^{\mathrm{T}} J \begin{bmatrix} I & \cdots & & 0 \\ \vdots & \ddots & & \vdots \\ 0 & & I & 0 \\ 0 & \cdots & 0 & P \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 & A_1P \\ \vdots & \ddots & & \vdots \\ 0 & & 0 & A_nP \\ 0 & \cdots & 0 & \Lambda \end{bmatrix} \tag{6.32}$$

proving that the spectrum of J lies in $[0,1)$. Since the Jacobian in (6.28) is arbitrarily well approximated by J when τ is sufficiently large, the spectral radius of the Jacobian must be less than 1 for τ sufficiently large. ∎

Remark 1 *Computations demonstrate that the iteration (3.6) – (3.7) converges to a minimizer of M for all $\tau > 0$ and even when the condition $\boldsymbol{\mu}^\star \notin \{Y\hat{\boldsymbol{e}}_j\}_{j=1}^n$ is violated. Furthermore, while the uniqueness of the minimizer is not guaranteed when the non-colinearity condition is violated, the iteration is found to converge to the median shown in Lemma 1 when τ is sufficiently small.*

7 Convergence of the Iterative Scheme for ℓ_1 PCA

The analysis of the scheme (4.10) – (4.11) begins with establishing the existence of a minimizer for H_k in (4.8). Recall the assumption in Section 4 that $S_1 = \mathcal{R}(Y_{\mathrm{c}}) = \mathbb{R}^m$ so that $S_k = \mathcal{R}(Y_k) = \mathcal{R}(V_{k-1})^\perp$, $k = 2, \ldots, m$, as seen in (4.3). Also, define

$$\mathbb{S}_k = \{\hat{\boldsymbol{v}} \in S_k : \|\hat{\boldsymbol{v}}\|_{\ell_2} = 1\}. \tag{7.1}$$

Lemma 4 *For H_k in (4.8) there exists a minimizer $\hat{\boldsymbol{v}}^\star$ over S_k which satisfies $\hat{\boldsymbol{v}}^\star \in \mathbb{S}_k$.*

Proof: Because of the properties of projections, it follows that $\|(\hat{\boldsymbol{v}}\hat{\boldsymbol{v}}^{\mathrm{T}} - I)Y_k\hat{\boldsymbol{e}}_j\|_{\ell_2} \leq \|Y_k\hat{\boldsymbol{e}}_j\|_{\ell_2}$ holds $\forall \hat{\boldsymbol{v}} \in \mathbb{S}_k$ and $\forall j$. Thus, by (4.8), $H_k(\boldsymbol{v}) \leq H_k(\boldsymbol{0})$ holds $\forall \boldsymbol{v} \in S_k$. With (4.8) it follows that H_k is not minimized in $\boldsymbol{v} = \boldsymbol{0}$ if it can be shown that for some $\hat{\boldsymbol{v}} \in \mathbb{S}_k$ and some j,

$$\|(\hat{\boldsymbol{v}}\hat{\boldsymbol{v}}^{\mathrm{T}} - I)Y_k\hat{\boldsymbol{e}}_j\|_{\ell_2} < \|Y_k\hat{\boldsymbol{e}}_j\|_{\ell_2}. \tag{7.2}$$

If no such j and $\hat{\boldsymbol{v}}$ were to exist, then the contrapositive of (7.2) and the properties of projections give

$$\|\hat{\boldsymbol{v}}\hat{\boldsymbol{v}}^{\mathrm{T}}Y_k\hat{\boldsymbol{e}}_j\|_{\ell_2}^2 = \|Y_k\hat{\boldsymbol{e}}_j\|_{\ell_2}^2 - \|(\hat{\boldsymbol{v}}\hat{\boldsymbol{v}}^{\mathrm{T}} - I)Y_k\hat{\boldsymbol{e}}_j\|_{\ell_2}^2 = 0, \quad \forall j \in \{1, \ldots, n\}, \quad \forall \hat{\boldsymbol{v}} \in \mathbb{S}_k. \tag{7.3}$$

By (4.2), (4.3) and (7.3), $\hat{\boldsymbol{v}}\hat{\boldsymbol{v}}^{\mathrm{T}}Y_k = \hat{\boldsymbol{v}}\hat{\boldsymbol{v}}^{\mathrm{T}}(I - V_{k-1}V_{k-1}^{\mathrm{T}})Y_{\mathrm{c}} = \hat{\boldsymbol{v}}\hat{\boldsymbol{v}}^{\mathrm{T}}Y_{\mathrm{c}} = 0$. Multiplying this result by $\hat{\boldsymbol{v}} \neq \boldsymbol{0}$ gives $\boldsymbol{0} = (\hat{\boldsymbol{v}}^{\mathrm{T}}\hat{\boldsymbol{v}})_{=1}\hat{\boldsymbol{v}}^{\mathrm{T}}Y_{\mathrm{c}} = \hat{\boldsymbol{v}}^{\mathrm{T}}Y_{\mathrm{c}}$. Yet $\hat{\boldsymbol{v}}^{\mathrm{T}}Y_{\mathrm{c}} = 0$ contradicts the consequence of (4.3) that the kernel of $Y_{\mathrm{c}}^{\mathrm{T}}$ or $\mathcal{R}(Y_{\mathrm{c}})^\perp$ is empty. Thus, H_k is not minimized at $\boldsymbol{v} = \boldsymbol{0}$. According to (4.8), H_l is constant along rays outside the origin. Therefore, the minimization can as well be carried out over $\mathbb{S}_k$. The claim follows since $\mathbb{S}_k$ is compact and H_k is continuous on $\mathbb{S}_k$. ∎

Lemma 5 *The first-order necessary optimality condition for a minimizer $\hat{\boldsymbol{v}}^\star \in \mathbb{S}_k$ of H_k over S_k given by Lemma 4 is that there exists $D^\star \in \mathbb{R}^{m\times n}$ satisfying*

$$\begin{gathered}(\hat{\boldsymbol{v}}^\star\hat{\boldsymbol{v}}^{\star\mathrm{T}} - I)Y_k\hat{\boldsymbol{e}}_j = \|(\hat{\boldsymbol{v}}^\star\hat{\boldsymbol{v}}^{\star\mathrm{T}} - I)Y_k\hat{\boldsymbol{e}}_j\|_{\ell_2}D^\star\hat{\boldsymbol{e}}_j, \quad \hat{\boldsymbol{v}}^{\star\mathrm{T}}D^\star\hat{\boldsymbol{e}}_j = 0, \quad \|D^\star\hat{\boldsymbol{e}}_j\|_{\ell_2} \leq 1, \\ j = 1, \ldots, n \\ \sum_{j=1}^n (\hat{\boldsymbol{v}}^{\star\mathrm{T}}Y_k\hat{\boldsymbol{e}}_j)D^\star\hat{\boldsymbol{e}}_j = \boldsymbol{0}.\end{gathered} \tag{7.4}$$

Proof: The necessary optimality condition for a minimizer $\boldsymbol{v}^\star$ is that $0 \in \partial H_k(\boldsymbol{v}^\star)$. By the chain-rule (see, e.g., [2], p. 233), the subdifferential of H_k is given by the sum of the respective subdifferentials,

$$\partial H_k(\boldsymbol{v}) = \sum_{j=1}^{n} \partial \left\| \left(\frac{\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}}{\|\boldsymbol{v}\|_{\ell_2}^2} - I \right) Y_k \hat{\boldsymbol{e}}_j \right\|_{\ell_2}, \quad \boldsymbol{v} \neq \boldsymbol{0}. \tag{7.5}$$

Thus, there exist $\boldsymbol{b}_j^\star \in \partial_{\boldsymbol{v}} \|(\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}/\|\boldsymbol{v}\|_{\ell_2}^2 - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}|_{\boldsymbol{v}=\boldsymbol{v}^\star}$, $j = 1, \ldots, n$, satisfying

$$\sum_{j=1}^{n} \boldsymbol{b}_j^\star = 0, \tag{7.6}$$

The respective subdifferentials are given according to

$$\partial_{\boldsymbol{v}} \left\| \left(\frac{\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}}{\|\boldsymbol{v}\|_{\ell_2}^2} - I \right) Y_k \hat{\boldsymbol{e}}_j \right\|_{\ell_2} = \left[\partial_{\boldsymbol{v}} \left(\frac{\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}}{\|\boldsymbol{v}\|_{\ell_2}^2} Y_k \hat{\boldsymbol{e}}_j \right) \right]^{\mathrm{T}} \underbrace{\partial_{\boldsymbol{w}} \|\boldsymbol{w} - Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}}_{\boldsymbol{w} = \boldsymbol{v}\boldsymbol{v}^{\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j / \|\boldsymbol{v}\|_{\ell_2}^2} \tag{7.7}$$

where

$$\partial_{\boldsymbol{v}} \left(\frac{\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}}{\|\boldsymbol{v}\|_{\ell_2}^2} Y_k \hat{\boldsymbol{e}}_j \right) = \frac{\boldsymbol{v}^{\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j}{\|\boldsymbol{v}\|_{\ell_2}^2} I + \frac{\boldsymbol{v}\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_k^{\mathrm{T}}}{\|\boldsymbol{v}\|_{\ell_2}^2} - \frac{2\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}}{\|\boldsymbol{v}\|_{\ell_2}^4} \boldsymbol{v}^{\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j \tag{7.8}$$

and

$$\partial_{\boldsymbol{w}} \|\boldsymbol{w} - Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2} = \begin{cases} \dfrac{\boldsymbol{w} - Y_k \hat{\boldsymbol{e}}_j}{\|\boldsymbol{w} - Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}}, & \boldsymbol{w} \neq Y_k \hat{\boldsymbol{e}}_j \\ B(0,1), & \boldsymbol{w} = Y_k \hat{\boldsymbol{e}}_j \end{cases} \tag{7.9}$$

with the unit ball $B(0,1)$. Let $\boldsymbol{c}_j^\star$ be chosen so that

$$\boldsymbol{c}_j^\star \in \begin{cases} \dfrac{(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j}{\|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}}, & Y_k \hat{\boldsymbol{e}}_j \neq \hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j \\ B(0,1), & Y_k \hat{\boldsymbol{e}}_j = \hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j \end{cases} \tag{7.10}$$

and

$$\boldsymbol{b}_j^\star = [\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j I + \hat{\boldsymbol{v}}^\star \hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_k^{\mathrm{T}} - 2\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} \hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j] \boldsymbol{c}_j^\star. \tag{7.11}$$

According to (7.10),

$$\|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2} \boldsymbol{c}_j^\star = (\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j, \quad j = 1, \ldots, n. \tag{7.12}$$

With (7.11) it follows for $Y_k \hat{\boldsymbol{e}}_j \neq \hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j$,

$$\begin{aligned} \boldsymbol{b}_j^\star &= \left[\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j I + Y_k \hat{\boldsymbol{e}}_j \hat{\boldsymbol{v}}^{\star\mathrm{T}} - 2\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j \hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} \right] \frac{(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j}{\|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} \\ &= (\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j) \frac{(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j}{\|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} = (\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j)(I - \hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}}) \frac{(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j}{\|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} \\ &= (\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j)(I - \hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}}) \boldsymbol{c}_j^\star \end{aligned} \tag{7.13}$$

and for $Y_k \hat{\boldsymbol{e}}_j = \hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j$,

$$\boldsymbol{b}_j^\star = [\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j I + \hat{\boldsymbol{v}}^\star \hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_k^{\mathrm{T}} - 2\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} \hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j] \boldsymbol{c}_j^\star = (\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j)(I - \hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}}) \boldsymbol{c}_j^\star. \tag{7.14}$$

Define

$$\boldsymbol{d}_j^\star = (I - \hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}}) \boldsymbol{c}_j^\star. \tag{7.15}$$

Combining (7.6), (7.13), (7.14) and (7.15) gives

$$0 = \sum_{j=1}^{n} \boldsymbol{b}_j^\star = \sum_{j=1}^{n} (\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j) \boldsymbol{d}_j^\star. \tag{7.16}$$

The claim (7.4) follows with $D^\star = \{\boldsymbol{d}_1^\star, \ldots, \boldsymbol{d}_n^\star\}$. ∎

Turning to the iteration (4.10) – (4.11) we observe that if convergence to a fixed point $\{D^\star, \hat{\boldsymbol{v}}^\star\}$ with $\|\hat{\boldsymbol{v}}^\star\|_{\ell_2} = 1$ can be guaranteed, then from (4.10) we have that $\hat{\boldsymbol{v}}^{\star\mathrm{T}} D_j^\star \hat{\boldsymbol{e}}_j = 0$ and

$$\|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2} D^\star \hat{\boldsymbol{e}}_j = (\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j, \quad j = 1, \ldots, n. \tag{7.17}$$

According to (4.11), the fixed point $\hat{\boldsymbol{v}}^\star$ satisfies

$$\hat{\boldsymbol{v}}^\star \|\boldsymbol{v}^\star\|_{\ell_2} = \boldsymbol{v}^\star = \hat{\boldsymbol{v}}^\star - \rho \sum_{j=1}^{n} \frac{(\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j)(\tau Y_k \hat{\boldsymbol{e}}_j - D_l^\star \hat{\boldsymbol{e}}_j)}{1 + \tau \|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} \tag{7.18}$$

where $\boldsymbol{v}^\star$ is defined by the right side of (7.18). Applying $(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I)$ to both sides of (7.18) gives

$$0 = (\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) \sum_{j=1}^{n} \frac{(\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j)(\tau Y_k \hat{\boldsymbol{e}}_j - D_l^\star \hat{\boldsymbol{e}}_j)}{1 + \tau \|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} = \sum_{j=1}^{n} (\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j) \frac{D_l^\star \hat{\boldsymbol{e}}_j + \tau (\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j}{1 + \tau \|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}}. \tag{7.19}$$

Combining (7.17) and (7.19) gives

$$0 = \sum_{j=1}^{n} (\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j) \frac{D^\star \hat{\boldsymbol{e}}_j + \tau \|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2} D^\star \hat{\boldsymbol{e}}_j}{1 + \tau \|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} = \sum_{j=1}^{n} (\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j) D^\star \hat{\boldsymbol{e}}_j. \tag{7.20}$$

Moreover, if $\|D_0 \hat{\boldsymbol{e}}_j\|_{\ell_2} \leq 1, j = 1, \ldots, n$, then the iterates $\{D_l\}$ also satisfy this bound, and hence $\|D^\star \hat{\boldsymbol{e}}_j\|_{\ell_2} \leq 1, j = 1, \ldots, n$. Thus $\{D^\star, \hat{\boldsymbol{v}}^\star\}$ must satisfy the necessary optimality condition (7.4).

Theorem 2 *Let $\{D^\star, \hat{\boldsymbol{v}}^\star\}$ satisfy (7.4) with $\hat{\boldsymbol{v}}^\star \in \mathbb{S}_k$ and suppose*

$$\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j \neq 0, \quad (\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j \neq \mathbf{0}, \quad j = 1, \ldots, n. \tag{7.21}$$

Then $\{D^\star, \hat{\boldsymbol{v}}^\star\}$ is a fixed point of the iteration (4.10) – (4.11) which is locally asymptotically stable for τ sufficiently large and for ρ sufficiently small.

Proof: Using (7.4) and substituting $D_l = D^\star$ and $\hat{\mathbf{v}}_l = \hat{\boldsymbol{v}}^\star$ on the right side of (4.10),

$$D_{l+1} \hat{\boldsymbol{e}}_j = \frac{(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I)(\tau Y_k \hat{\boldsymbol{e}}_j - D^\star \hat{\boldsymbol{e}}_j)}{1 + \tau \|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} = \frac{\tau \|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2} D^\star \hat{\boldsymbol{e}}_j + D^\star \hat{\boldsymbol{e}}_j}{1 + \tau \|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} = D^\star \hat{\boldsymbol{e}}_j \tag{7.22}$$

gives $D_{l+1} = D^\star$. Also by (7.4),

$$\begin{aligned} 0 &= \sum_{j=1}^{n} (\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j) D^\star \hat{\boldsymbol{e}}_j \frac{1 + \tau \|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}}{1 + \tau \|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} \\ &= \sum_{j=1}^{n} (\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j) \frac{D^\star \hat{\boldsymbol{e}}_j + \tau (\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j}{1 + \tau \|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} \\ &= (\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) \sum_{j=1}^{n} (\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j) \frac{\tau Y_k \hat{\boldsymbol{e}}_j - D^\star \hat{\boldsymbol{e}}_j}{1 + \tau \|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} \end{aligned} \tag{7.23}$$

and hence

$$\boldsymbol{v}^\star = \hat{\boldsymbol{v}}^\star - \rho \sum_{j=1}^{n} \frac{(\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_l \hat{\boldsymbol{e}}_j)(\tau Y_k \hat{\boldsymbol{e}}_j - D^\star \hat{\boldsymbol{e}}_j)}{1 + \tau \|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} = \hat{\boldsymbol{v}}^\star \left[1 - \rho \hat{\boldsymbol{v}}^{\star\mathrm{T}} \sum_{j=1}^{n} \frac{(\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_l \hat{\boldsymbol{e}}_j)(\tau Y_k \hat{\boldsymbol{e}}_j - D^\star \hat{\boldsymbol{e}}_j)}{1 + \tau \|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} \right] \tag{7.24}$$

satisfies $\boldsymbol{v}^\star = \hat{\boldsymbol{v}}^\star \|\boldsymbol{v}^\star\|_{\ell_2}$. Thus, setting $D_l = D^\star$ and $\hat{\mathbf{v}}_l = \hat{\boldsymbol{v}}^\star$ on the right side of (4.11) gives $\hat{\mathbf{v}}_{l+1} = \hat{\boldsymbol{v}}^\star$. Therefore, $\{D^\star, \hat{\boldsymbol{v}}^\star\}$ is a fixed point of the iteration (4.10) – (4.11).

To establish the stability of the fixed point, define

$$\boldsymbol{F}_j(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_n, \boldsymbol{v}) = \frac{(\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}} - I)(\tau Y_k \hat{\boldsymbol{e}}_j - \boldsymbol{d}_j)}{1 + \tau \|(\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}}, \quad j = 1, \ldots, n \tag{7.25}$$

and

$$\begin{aligned} \boldsymbol{G}(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_n, \boldsymbol{v}) &= \frac{\boldsymbol{g}(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_n, \boldsymbol{v})}{\|\boldsymbol{g}(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_n, \boldsymbol{v})\|_{\ell_2}} \\ \boldsymbol{g}(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_n, \boldsymbol{v}) &= \boldsymbol{v} - \rho \sum_{j=1}^{n} \frac{(\boldsymbol{v}^{\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j)(\tau Y_k \hat{\boldsymbol{e}}_j - \boldsymbol{d}_j)/\|\boldsymbol{v}\|_{\ell_2}}{1 + \tau \|(\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} \end{aligned} \tag{7.26}$$

so that (4.10) – (4.11) is given by

$$D_{l+1}\hat{\boldsymbol{e}}_j = \boldsymbol{F}_j(D_l \hat{\boldsymbol{e}}_1, \ldots, D_l \hat{\boldsymbol{e}}_n, \hat{\mathbf{v}}_l), \quad j = 1, \ldots, n, \quad \hat{\mathbf{v}}_{l+1} = \boldsymbol{G}(D_{l+1}\hat{\boldsymbol{e}}_1, \ldots, D_{l+1}\hat{\boldsymbol{e}}_n, \hat{\mathbf{v}}_l). \tag{7.27}$$

The claimed stability will follow once it is shown that the Jacobian of this mapping from the $(n+1)$-fold Cartesian product $(S_k)^{n+1}$ into $(S_k)^{n+1}$ evaluated at $\{D^\star, \hat{\boldsymbol{v}}^\star\} = \{\boldsymbol{d}_1^\star, \ldots, \boldsymbol{d}_n^\star, \hat{\boldsymbol{v}}^\star\}$ has only eigenvalues with magnitude less than 1 when τ is sufficiently large and ρ is sufficiently small. For (7.25),

$$\frac{\partial \boldsymbol{F}_j}{\partial \boldsymbol{d}_i}(\boldsymbol{d}_1^\star, \ldots, \boldsymbol{d}_n^\star, \hat{\boldsymbol{v}}^\star) = \frac{I - \hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}}}{1 + \tau \|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} \overset{\tau \to \infty}{\longrightarrow} 0 \tag{7.28}$$

and

$$\begin{aligned} \frac{\partial \boldsymbol{F}_j}{\partial \boldsymbol{v}}(\boldsymbol{d}_1^\star, \ldots, \boldsymbol{d}_n^\star, \boldsymbol{v}) = \Bigg\{ & \frac{\boldsymbol{v}^{\mathrm{T}}(\tau Y_k \hat{\boldsymbol{e}}_j - \boldsymbol{d}_j^\star) I + \boldsymbol{v}(\tau Y_k \hat{\boldsymbol{e}}_j - \boldsymbol{d}_j^\star)^{\mathrm{T}}}{1 + \tau \|(\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} \\ & - \frac{(\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}} - I)(\tau Y_k \hat{\boldsymbol{e}}_j - \boldsymbol{d}_j)}{[1 + \tau \|(\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}]^2} \left[\left(\boldsymbol{v}^{\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j I + \boldsymbol{v} \hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_k^{\mathrm{T}} \right)^{\mathrm{T}} \frac{\tau (\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j}{\|(\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} \right]^{\mathrm{T}} \Bigg\}_{\boldsymbol{v} = \hat{\boldsymbol{v}}^\star} \\ \overset{\tau \to \infty}{\longrightarrow} & \frac{\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j I + \hat{\boldsymbol{v}}^\star \hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_k^{\mathrm{T}}}{\|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}} - (\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_j) \frac{(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j \hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_k^{\mathrm{T}} (\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I)}{\|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_j\|_{\ell_2}^3} := A_j \end{aligned} \tag{7.29}$$

For (7.26),

$$\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{d}_i}(\boldsymbol{d}_1^\star, \ldots, \boldsymbol{d}_n^\star, \hat{\boldsymbol{v}}^\star) = \rho \frac{(\hat{\boldsymbol{v}}^{\star\mathrm{T}} Y_k \hat{\boldsymbol{e}}_i) I / \|\hat{\boldsymbol{v}}^\star\|_{\ell_2}}{1 + \tau \|(\hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}} - I) Y_k \hat{\boldsymbol{e}}_i\|_{\ell_2}} \overset{\tau \to \infty}{\longrightarrow} 0 \tag{7.30}$$

and with $\hat{\boldsymbol{g}}^\star = \boldsymbol{g}(\boldsymbol{d}_1^\star, \ldots, \boldsymbol{d}_n^\star, \hat{\boldsymbol{v}}^\star) = \hat{\boldsymbol{v}}^\star$,

$$\begin{aligned} \frac{\partial \boldsymbol{G}}{\partial \boldsymbol{d}_i}(\boldsymbol{d}_1^\star, \ldots, \boldsymbol{d}_n^\star, \hat{\boldsymbol{v}}^\star) &= \frac{1}{\|\hat{\boldsymbol{g}}^\star\|_{\ell_2}} \left(I - \frac{\hat{\boldsymbol{g}}^\star \hat{\boldsymbol{g}}^{\star\mathrm{T}}}{\|\hat{\boldsymbol{g}}^\star\|_{\ell_2}} \right) \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{d}_i}(\boldsymbol{d}_1^\star, \ldots, \boldsymbol{d}_n^\star, \hat{\boldsymbol{v}}^\star) \\ &= (I - \hat{\boldsymbol{v}}^\star \hat{\boldsymbol{v}}^{\star\mathrm{T}}) \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{d}_i}(\boldsymbol{d}_1^\star, \ldots, \boldsymbol{d}_n^\star, \hat{\boldsymbol{v}}^\star) \overset{\tau \to \infty}{\longrightarrow} 0. \end{aligned} \tag{7.31}$$

Also

$$\begin{aligned}
\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{v}}(\boldsymbol{d}_1^\star,\dots,\boldsymbol{d}_n^\star,\boldsymbol{v}) &= \Bigg\{ I-\rho\sum_{j=1}^n \frac{(\tau Y_k\hat{\boldsymbol{e}}_j-\boldsymbol{d}_j^\star)\hat{\boldsymbol{e}}_j^{\mathrm{T}}Y_k^{\mathrm{T}}}{1+\tau\|(\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}-I)\|_{\ell_2}}\left(I-\frac{\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}}{\|\boldsymbol{v}\|_{\ell_2}^2}\right)\frac{1}{\|\boldsymbol{v}\|_{\ell_2}} \\
&+\rho\sum_{j=1}^n \frac{(\boldsymbol{v}^{\mathrm{T}}Y_k\hat{\boldsymbol{e}}_j)(\tau Y_k\hat{\boldsymbol{e}}_j-\boldsymbol{d}_j)/\|\boldsymbol{v}\|_{\ell_2}}{[1+\tau\|(\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}-I)Y_k\hat{\boldsymbol{e}}_j\|_{\ell_2}]^2}\left[\left(\boldsymbol{v}^{\mathrm{T}}Y_k\hat{\boldsymbol{e}}_j I+\boldsymbol{v}\hat{\boldsymbol{e}}_j^{\mathrm{T}}Y_k^{\mathrm{T}}\right)^{\mathrm{T}}\frac{\tau(\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}-I)Y_k\hat{\boldsymbol{e}}_j}{\|(\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}-I)Y_k\hat{\boldsymbol{e}}_j\|_{\ell_2}}\right]^{\mathrm{T}}\Bigg\}_{\boldsymbol{v}=\hat{\boldsymbol{v}}^\star} \\
&= I-\rho\sum_{j=1}^n \frac{(\tau Y_k\hat{\boldsymbol{e}}_j-\boldsymbol{d}_j^\star)\hat{\boldsymbol{e}}_j^{\mathrm{T}}Y_k^{\mathrm{T}}}{1+\tau\|(\hat{\boldsymbol{v}}^\star\hat{\boldsymbol{v}}^{\star\mathrm{T}}-I)Y_k\hat{\boldsymbol{e}}_j\|_{\ell_2}}\left[1+\frac{\tau(\hat{\boldsymbol{v}}^{\mathrm{T}}Y_k\hat{\boldsymbol{e}}_j)^2}{[1+\tau\|(\hat{\boldsymbol{v}}\hat{\boldsymbol{v}}^{\mathrm{T}}-I)Y_k\hat{\boldsymbol{e}}_j\|_{\ell_2}]\|(\hat{\boldsymbol{v}}\hat{\boldsymbol{v}}^{\mathrm{T}}-I)Y_k\hat{\boldsymbol{e}}_j\|_{\ell_2}}\right]\times \\
&\qquad\qquad (I-\hat{\boldsymbol{v}}^\star\hat{\boldsymbol{v}}^{\star\mathrm{T}}) \\
&\overset{\tau\to\infty}{\longrightarrow} I-\rho\left[\sum_{j=1}^n(\hat{\boldsymbol{v}}^{\star\mathrm{T}}Y_k\hat{\boldsymbol{e}}_j)^2\frac{Y_k\hat{\boldsymbol{e}}_j\hat{\boldsymbol{e}}_j^{\mathrm{T}}Y_k^{\mathrm{T}}}{\|(\hat{\boldsymbol{v}}^\star\hat{\boldsymbol{v}}^{\star\mathrm{T}}-I)Y_k\hat{\boldsymbol{e}}_j\|_{\ell_2}^3}\right](I-\hat{\boldsymbol{v}}^\star\hat{\boldsymbol{v}}^{\star\mathrm{T}})
\end{aligned} \tag{7.32}$$

and

$$\begin{aligned}
\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{v}}(\boldsymbol{d}_1^\star,\dots,\boldsymbol{d}_n^\star,\hat{\boldsymbol{v}}^\star) &= \frac{1}{\|\boldsymbol{g}^\star\|_{\ell_2}}\left(I-\frac{\boldsymbol{g}^\star\boldsymbol{g}^{\star\mathrm{T}}}{\|\boldsymbol{g}^\star\|_{\ell_2}}\right)\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{v}}(\boldsymbol{d}_1^\star,\dots,\boldsymbol{d}_n^\star,\hat{\boldsymbol{v}}^\star) \\
&= (I-\hat{\boldsymbol{v}}^\star\hat{\boldsymbol{v}}^{\star\mathrm{T}})\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{v}}(\boldsymbol{d}_1^\star,\dots,\boldsymbol{d}_n^\star,\hat{\boldsymbol{v}}^\star) \\
\overset{\tau\to\infty}{\longrightarrow} (I-\hat{\boldsymbol{v}}^\star\hat{\boldsymbol{v}}^{\star\mathrm{T}})&\left[I-\rho\sum_{j=1}^n(\hat{\boldsymbol{v}}^{\star\mathrm{T}}Y_k\hat{\boldsymbol{e}}_j)^2\frac{Y_k\hat{\boldsymbol{e}}_j\hat{\boldsymbol{e}}_j^{\mathrm{T}}Y_k^{\mathrm{T}}}{\|(\hat{\boldsymbol{v}}^\star\hat{\boldsymbol{v}}^{\star\mathrm{T}}-I)Y_k\hat{\boldsymbol{e}}_j\|_{\ell_2}^3}\right](I-\hat{\boldsymbol{v}}^\star\hat{\boldsymbol{v}}^{\star\mathrm{T}}) =: B
\end{aligned} \tag{7.33}$$

Thus, the Jacobian satisfies

$$\frac{\partial(\boldsymbol{F}_1,\dots,\boldsymbol{F}_n,\boldsymbol{G})}{\partial(\boldsymbol{d}_1,\dots,\boldsymbol{d}_n,\boldsymbol{\mu})}(\boldsymbol{d}_1^\star,\dots,\boldsymbol{d}_n^\star,\boldsymbol{\mu}^\star)\overset{\tau\to\infty}{\longrightarrow}\begin{bmatrix} 0 & \cdots & 0 & A_1 \\ \vdots & \ddots & & \vdots \\ 0 & & 0 & A_n \\ 0 & \cdots & 0 & B \end{bmatrix} =: J. \tag{7.34}$$

By the definition $S_k=\mathcal{R}(Y_k)$ prior to (4.3), it follows from (7.29) that $A_j: S_k\to S_k$, $j=1,\dots,n$, and from (7.33) that $B: S_k\to S_k$. In the same way, it follows from (7.34) that $J:(S_k)^{n+1}\to(S_k)^{n+1}$. It will be shown that $B: S_k\to S_k$ has spectrum in $(-1,1)$. Let B be expressed as $B=(I-\hat{\boldsymbol{v}}^\star\hat{\boldsymbol{v}}^{\star\mathrm{T}})-\rho\hat{C}$ where

$$\hat{C}=(I-\hat{\boldsymbol{v}}^\star\hat{\boldsymbol{v}}^{\star\mathrm{T}})C(I-\hat{\boldsymbol{v}}^\star\hat{\boldsymbol{v}}^{\star\mathrm{T}}),\quad C=\sum_{j=1}^n(\hat{\boldsymbol{v}}^{\star\mathrm{T}}Y_k\hat{\boldsymbol{e}}_j)^2\frac{Y_k\hat{\boldsymbol{e}}_j\hat{\boldsymbol{e}}_j^{\mathrm{T}}Y_k^{\mathrm{T}}}{\|(\hat{\boldsymbol{v}}^\star\hat{\boldsymbol{v}}^{\star\mathrm{T}}-I)Y_k\hat{\boldsymbol{e}}_j\|_{\ell_2}^3}. \tag{7.35}$$

Suppose there were a $\boldsymbol{v}\in S_k$ for which $\boldsymbol{v}^{\mathrm{T}}C\boldsymbol{v}=0$. Then by (4.2) and (4.3), $\boldsymbol{v}^{\mathrm{T}}Y_k=\boldsymbol{v}^{\mathrm{T}}(I-V_{k-1}V_{k-1}^{\mathrm{T}})Y_{\mathrm{c}}=\boldsymbol{v}^{\mathrm{T}}Y_{\mathrm{c}}$ and with (7.21),

$$0=\sum_{j=1}^n(\hat{\boldsymbol{v}}^{\star\mathrm{T}}Y_k\hat{\boldsymbol{e}}_j)^2\frac{|\boldsymbol{v}^{\mathrm{T}}Y_{\mathrm{c}}\hat{\boldsymbol{e}}_j|^2}{\|(\hat{\boldsymbol{v}}^\star\hat{\boldsymbol{v}}^{\star\mathrm{T}}-I)Y_k\hat{\boldsymbol{e}}_j\|}\geq\|\boldsymbol{v}^{\mathrm{T}}Y_{\mathrm{c}}\|_{\ell_2}^2\left[\min_{1\leq j\leq n}\frac{(\hat{\boldsymbol{v}}^{\star\mathrm{T}}Y_k\hat{\boldsymbol{e}}_j)^2}{\|(\hat{\boldsymbol{v}}^\star\hat{\boldsymbol{v}}^{\star\mathrm{T}}-I)Y_k\hat{\boldsymbol{e}}_j\|}\right]_{>0}, \tag{7.36}$$

and $\boldsymbol{v}^{\mathrm{T}}Y_{\mathrm{c}}=0$ would contradict the assumption that $S_1=\mathcal{R}(Y_{\mathrm{c}})=\mathbb{R}^m$. Thus, there exist $0<\lambda_{\min}\leq\lambda_{\max}$ such that C satisfies

$$\lambda_{\min}\|\boldsymbol{x}\|_{\ell_2}^2\leq\boldsymbol{x}^{\mathrm{T}}C\boldsymbol{x}\leq\lambda_{\max}\|\boldsymbol{x}\|_{\ell_2}^2,\quad\forall\boldsymbol{x}\in S_k. \tag{7.37}$$

Let $S_k=U_k\oplus W_k$ where $W_k=\mathrm{span}\{\hat{\boldsymbol{v}}^\star\}$. Clearly, $\lambda=0$ is an eigenvalue of B associated with the eigenvector $\hat{\boldsymbol{v}}^\star\in W_k$. Eigenvalues for eigenvectors in U_k will now be estimated. Let ρ be small enough

that $-1 < 1-\rho\lambda_{\max} < 1-\rho\lambda_{\min} < 1$. Then for $\boldsymbol{x} \in U_k$ it follows with the definition of $\hat{C}$ and C that $\boldsymbol{x}^{\mathrm{T}} B \boldsymbol{x} = \|\boldsymbol{x}\|_{\ell_2}^2 - \rho \boldsymbol{x}^{\mathrm{T}} \hat{C} \boldsymbol{x} = \|\boldsymbol{x}\|_{\ell_2}^2 - \rho \boldsymbol{x}^{\mathrm{T}} C \boldsymbol{x}$ and

$$-\|\boldsymbol{x}\|_{\ell_2}^2 < (1-\rho\lambda_{\max})\|\boldsymbol{x}\|_{\ell_2}^2 \le \|\boldsymbol{x}\|_{\ell_2}^2 - \rho \boldsymbol{x}^{\mathrm{T}} C \boldsymbol{x} \le (1-\rho\lambda_{\min})\|\boldsymbol{x}\|_{\ell_2}^2 < \|\boldsymbol{x}\|_{\ell_2}^2. \tag{7.38}$$

Thus, the spectrum of $B : S_k \to S_k$ lies in $(-1,1)$. Recalling that the dimension of S_k is $m-k+1$, let $P \in \mathbb{R}^{m\times(m-k+1)}$ be chosen with orthonormal columns such that $P^{\mathrm{T}} B P = \Lambda = \mathrm{diag}\{\lambda_i\}_{i=1}^{m-k+1}$ where $\lambda_i \in (-1,1)$. Also set $Z \in \mathbb{R}^{m\times(m-k+1)}$ with all zero entries. Further let $I \in \mathbb{R}^{m\times m}$ and $0 \in \mathbb{R}^{m\times m}$ in (7.39) below denote the identity and the zero matrix respectively. Then the matrix J satisfies

$$\begin{bmatrix} I & \cdots & & Z \\ \vdots & \ddots & & \vdots \\ 0 & & I & Z \\ 0 & \cdots & 0 & P \end{bmatrix}^{\mathrm{T}} J \begin{bmatrix} I & \cdots & & Z \\ \vdots & \ddots & & \vdots \\ 0 & & I & Z \\ 0 & \cdots & 0 & P \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 & A_1 P \\ \vdots & \ddots & & \vdots \\ 0 & & 0 & A_n P \\ Z^{\mathrm{T}} Z & \cdots & Z^{\mathrm{T}} Z & \Lambda \end{bmatrix} \tag{7.39}$$

proving that the spectrum of J lies in $(-1,1)$. Since the Jacobian in (6.28) is arbitrarily well approximated by J when τ is sufficiently large, its spectrum lies strictly within the ball of radius 1 for τ sufficiently large and for ρ sufficiently small. ∎

Remark 2 *Computations demonstrate that the iteration* (4.10) – (4.11) *can converge to a minimizer for H_k even when the condition* (7.21) *is violated. Such a case is illustrated in Fig. 2. Variations of the example illustrated in Fig. 2, in which data points lie on the boundary of a rectangle instead of a square, indicate an advantage to having sphered the data by ℓ_2 means according to* (2.8) *and* (2.10) *before proceeding with the methods of Section 4.*

8 Convergence of the Iterative Scheme for ℓ_1 ICA

The analysis of the scheme (5.11) begins with establishing existence of a maximizer for G_l in (5.9). Recall the assumption in Section 5 that $T_1 = \mathcal{R}(Y_{\mathrm{s}}) = \mathbb{R}^m$ so that $T_l = \mathcal{R}(Y_l) = \mathcal{R}(U_{l-1}^{\mathrm{T}})^{\perp}$, $l = 2,\ldots,m$, as seen in (5.7). Also, define

$$\mathbb{T}_l = \{\hat{\boldsymbol{u}} \in T_l : \|\hat{\boldsymbol{u}}\|_{\ell_2} = 1\}. \tag{8.1}$$

Lemma 6 *For G_l in* (5.9) *there exists a maximizer $\hat{\boldsymbol{u}}^\star$ over T_l which satisfies $\hat{\boldsymbol{u}}^\star \in \mathbb{T}_l$.*

Proof: Choose a $\hat{\boldsymbol{u}} \in \mathbb{T}_l$ so that $G_l(\hat{\boldsymbol{u}}) \neq 0$. If no such $\hat{\boldsymbol{u}}$ were to exist, then by (2.16) and (5.7),

$$0 = \hat{\boldsymbol{u}}^{\mathrm{T}} Y_l = \hat{\boldsymbol{u}}^{\mathrm{T}} (I - U_{l-1}^{\mathrm{T}} U_{k-1}) Y_{\mathrm{s}} = \hat{\boldsymbol{u}}^{\mathrm{T}} Y_{\mathrm{s}} = 0 \tag{8.2}$$

would contradict the assumption that $T_1 = \mathcal{R}(Y_{\mathrm{s}}) = \mathbb{R}^m$. Thus, G_l is not maximized at $\boldsymbol{u} = \boldsymbol{0}$. According to (5.9), G_l is constant along rays outside the origin. Thus, the maximization can as well be carried out on $\mathbb{T}_l$. The claim follows since $\mathbb{T}_l$ is compact and G_l is continuous on $\mathbb{T}_l$. ∎

Lemma 7 *For any $\boldsymbol{u} \in T_l \backslash \{0\}$ the directional derivative of G_l in the direction of $\boldsymbol{w} \in T_l$ is given by*

$$\partial G_l(\boldsymbol{u};\boldsymbol{w}) = \frac{1}{\|\boldsymbol{u}\|_{\ell_2}} \left[\sum_{j=1}^{n} \delta_j(\boldsymbol{u},\boldsymbol{w}) Y_l \hat{\boldsymbol{e}}_j - \sum_{j=1}^{n} \frac{|\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u}|}{\|\boldsymbol{u}\|_{\ell_2}} \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|_{\ell_2}} \right]^{\mathrm{T}} \boldsymbol{w} \tag{8.3}$$

where

$$\delta_j(\boldsymbol{u},\boldsymbol{w}) = \begin{cases} \sigma(\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u}), & \hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u} \neq 0 \\ \sigma(\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{w}), & \hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u} = 0 \end{cases} \qquad \sigma(t) = \mathrm{sign}(t). \tag{8.4}$$

Proof: For $\boldsymbol{u} \in T_l \backslash \{0\}$, $\boldsymbol{w} \in T_l$ and $t > 0$ sufficiently small that $\boldsymbol{u} + t\boldsymbol{w} \neq 0$,

$$G_l(\boldsymbol{u}+t\boldsymbol{w}) - G_l(\boldsymbol{u}) = \sum_{j=1}^{n} \frac{|\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}}(\boldsymbol{u}+t\boldsymbol{w})| - |\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u}|}{\|\boldsymbol{u}+t\boldsymbol{w}\|_{\ell_2}} - \sum_{j=1}^{n} |\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u}| \frac{\|\boldsymbol{u}+t\boldsymbol{w}\|_{\ell_2} - \|\boldsymbol{u}\|_{\ell_2}}{\|\boldsymbol{u}+t\boldsymbol{w}\|_{\ell_2}\|\boldsymbol{u}\|_{\ell_2}}. \quad (8.5)$$

For $\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u} \neq 0$ the terms of the first sum in (8.5) satisfy

$$\begin{aligned} \lim_{t\to 0^+} \frac{|\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}}(\boldsymbol{u}+t\boldsymbol{w})| - |\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u}|}{t\|\boldsymbol{u}+t\boldsymbol{w}\|_{\ell_2}} &= \lim_{t\to 0^+} \frac{(2t\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u} + t^2 \hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u})(\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u})/t}{(|\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}}(\boldsymbol{u}+t\boldsymbol{w})| + |\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u}|)\|\boldsymbol{u}+t\boldsymbol{w}\|_{\ell_2}} \\ &= \sigma(\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u}) \frac{\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{w}}{\|\boldsymbol{u}\|_{\ell_2}} \end{aligned} \quad (8.6)$$

and for $\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u} = 0$,

$$\lim_{t\to 0^+} \frac{|\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}}(\boldsymbol{u}+t\boldsymbol{w})| - |\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u}|}{t\|\boldsymbol{u}+t\boldsymbol{w}\|_{\ell_2}} = \frac{|\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{w}|}{\|\boldsymbol{u}\|_{\ell_2}} = \sigma(\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{w}) \frac{\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{w}}{\|\boldsymbol{u}\|_{\ell_2}}. \quad (8.7)$$

The terms of the second sum in (8.5) satisfy

$$\lim_{t\to 0^+} \frac{\|\boldsymbol{u}+t\boldsymbol{w}\|_{\ell_2} - \|\boldsymbol{u}\|_{\ell_2}}{t\|\boldsymbol{u}+t\boldsymbol{w}\|_{\ell_2}\|\boldsymbol{u}\|_{\ell_2}} = \lim_{t\to 0^+} \frac{(2t\boldsymbol{u}^{\mathrm{T}}\boldsymbol{w} + t^2\|\boldsymbol{w}\|_{\ell_2}^2)/t}{\|\boldsymbol{u}+t\boldsymbol{w}\|_{\ell_2}\|\boldsymbol{u}\|_{\ell_2}(\|\boldsymbol{u}+t\boldsymbol{w}\|_{\ell_2} + \|\boldsymbol{u}\|_{\ell_2})} = \frac{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{w}}{\|\boldsymbol{u}\|_{\ell_2}^3}. \quad (8.8)$$

Combining these calculations gives (8.3). ∎

Lemma 7 is now used to prove Lemma 8. For the following, let $\sigma(\boldsymbol{v}) = \{\sigma(v_i)\}$ where $\boldsymbol{v} = \{v_i\}$ and $\sigma(t) = \mathrm{sign}(t)$.

Lemma 8 *The first-order necessary optimality condition for a maximizer $\hat{\boldsymbol{u}}^\star \in \mathbb{T}_l$ of G_l over T_l given by Lemma 6 is*

$$Y_l \sigma(Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star) = \|Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star\|_{\ell_1} \hat{\boldsymbol{u}}^\star \quad (8.9)$$

and the sets $\mathcal{Y} = \{j : Y_l \hat{\boldsymbol{e}}_j = \boldsymbol{0}\}$ and $\mathcal{S} = \{j : \boldsymbol{u}^{\star\mathrm{T}} Y_l \hat{\boldsymbol{e}}_j = 0\}$ agree. As a consequence, $\exists \epsilon > 0$ such that

$$\nabla G_l(\boldsymbol{u}) = \frac{1}{\|\boldsymbol{u}\|_{\ell_2}} \left[Y_l \sigma(Y_l^{\mathrm{T}} \boldsymbol{u}) - \frac{\|Y_l^{\mathrm{T}} \boldsymbol{u}\|_{\ell_1}}{\|\boldsymbol{u}\|_{\ell_2}} \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|_{\ell_2}} \right], \quad \forall \boldsymbol{u} \in B(\hat{\boldsymbol{u}}^\star, \epsilon). \quad (8.10)$$

Proof: Let $\hat{\boldsymbol{u}}^\star \in \mathbb{T}_l$ be a maximizer for G_l guaranteed by Lemma 6. By (8.3),

$$\partial G_l(\hat{\boldsymbol{u}}^\star; \boldsymbol{w}) = \lim_{t\to 0^+} \frac{G_l(\hat{\boldsymbol{u}}^\star + t\boldsymbol{w}) - G_l(\hat{\boldsymbol{u}}^\star)}{t} \leq 0, \quad \forall \boldsymbol{w} \in T_l. \quad (8.11)$$

By decomposing sums into indices in $\mathcal{S}$ and $\mathcal{S}^{\mathrm{c}}$,

$$\begin{aligned} &\partial G_l(\hat{\boldsymbol{u}}^\star; \boldsymbol{w}) \\ &= \left[\sum_{j\in\mathcal{S}} \sigma(\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{w}) Y_l \hat{\boldsymbol{e}}_j + \sum_{j\in\mathcal{S}^{\mathrm{c}}} \sigma(\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star) Y_l \hat{\boldsymbol{e}}_j - \sum_{j=1}^{n} |\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star| \hat{\boldsymbol{u}}^\star \right]^{\mathrm{T}} \boldsymbol{w} \\ &= \sum_{j\in\mathcal{S}} |\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{w}| + \left[\sum_{j\in\mathcal{S}^{\mathrm{c}}} \sigma(\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star) Y_l \hat{\boldsymbol{e}}_j + \sum_{j\in\mathcal{S}} \underbrace{\sigma(\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star)}_{=0} Y_l \hat{\boldsymbol{e}}_j - \sum_{j=1}^{n} |\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star| \hat{\boldsymbol{u}}^\star \right]^{\mathrm{T}} \boldsymbol{w} \\ &= \sum_{j\in\mathcal{S}} |\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{w}| + \boldsymbol{v}^{\star\mathrm{T}} \boldsymbol{w} \end{aligned} \quad (8.12)$$

where

$$\boldsymbol{v}^\star = \sum_{j=1}^{n} \sigma(\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star) Y_l \hat{\boldsymbol{e}}_j - \sum_{j=1}^{n} |\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star| \hat{\boldsymbol{u}}^\star = Y_l \sigma(Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star) - \|Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star\|_{\ell_1} \hat{\boldsymbol{u}}^\star. \tag{8.13}$$

Combining (8.11) and (8.12), it follows

$$\forall i \in \mathcal{S} \quad |\hat{\boldsymbol{e}}_i^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{w}| \le \sum_{j \in \mathcal{S}} |\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{w}| \le 0, \quad \forall \boldsymbol{w} \in \boldsymbol{v}^{\star\perp}. \tag{8.14}$$

This implies the existence of $\{\gamma_j\}_{j \in \mathcal{S}}$ (possibly zero) such that

$$Y_l \hat{\boldsymbol{e}}_j = \gamma_j \boldsymbol{v}^\star, \quad \forall j \in \mathcal{S}. \tag{8.15}$$

Now with $\boldsymbol{w} = \boldsymbol{v}^\star$, combining (8.11), (8.12) and (8.15) gives

$$0 \ge \sum_{j \in \mathcal{S}} |\gamma_j| \|\boldsymbol{v}^\star\|_{\ell_2}^2 + \|\boldsymbol{v}^\star\|_{\ell_2}^2 \tag{8.16}$$

or $\boldsymbol{v}^\star = \boldsymbol{0}$. With (8.13), the optimality condition (8.9) follows.

According to (8.15), $\mathcal{S} \subset \mathcal{Y}$ holds. Since $\mathcal{Y} \subset \mathcal{S}$ always holds, it follows that $\mathcal{Y} = \mathcal{S}$. Note that

$$G_l(\boldsymbol{u}) = \frac{1}{\|\boldsymbol{u}\|_{\ell_2}} \sum_{j=1}^{n} |\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u}| = \frac{1}{\|\boldsymbol{u}\|_{\ell_2}} \sum_{j \in \mathcal{Y}^{\mathrm{c}}} |\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u}| \tag{8.17}$$

is smooth for $\boldsymbol{u} \in B(\hat{\boldsymbol{u}}^\star, \epsilon)$ for some $\epsilon > 0$. As a result, G_l is smooth in $B(\hat{\boldsymbol{u}}^\star, \epsilon)$ and can be differentiated directly to obtain (8.10). ∎

Turning to the iteration (5.11) we observe that if convergence to a fixed point $\hat{\boldsymbol{u}}^\star$ can be guaranteed, then multiplying $\hat{\boldsymbol{u}}^{\star\mathrm{T}}$ by $\|\boldsymbol{u}^\star\| \hat{\boldsymbol{u}}^\star = \boldsymbol{u}^\star = \hat{\boldsymbol{u}}^\star + \tau[Y_l \sigma(Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star) - \hat{\boldsymbol{u}}^\star \|\hat{\boldsymbol{u}}^{\star\mathrm{T}} Y_l\|_{\ell_1}]$ gives

$$\|\boldsymbol{u}^\star\|_{\ell_2} = \hat{\boldsymbol{u}}^{\star\mathrm{T}} \boldsymbol{u}^\star = 1 + \tau[\hat{\boldsymbol{u}}^{\star\mathrm{T}} Y_l \sigma(Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star) - \|\hat{\boldsymbol{u}}^{\star\mathrm{T}} Y_l\|_{\ell_1}] = 1 \tag{8.18}$$

where the last equality follows with $(Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star)^{\mathrm{T}} \sigma(Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star) = \|\hat{\boldsymbol{u}}^{\star\mathrm{T}} Y_l\|_{\ell_1}$. Since $\|\boldsymbol{u}^\star\|_{\ell_2} = \|\hat{\boldsymbol{u}}^\star\|_{\ell_2} = 1$ holds, it follows that $\boldsymbol{u}^\star = \hat{\boldsymbol{u}}^\star$ and hence $\hat{\boldsymbol{u}}^\star$ satisfies (8.9).

Theorem 3 *Let $\hat{\boldsymbol{u}}^\star \in \mathbb{T}_l$ satisfy (8.9) where the set $\mathcal{S} = \{j : \hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star = 0\}$ is empty. Then $\hat{\boldsymbol{u}}^\star$ is a fixed point of the iteration (5.11) which is locally asymptotically stable for τ sufficiently small.*

Proof: Setting $\hat{\mathbf{u}}_{k+1} = \hat{\boldsymbol{u}}^\star$ on the right side of (5.11) and using (8.9) shows that $\mathbf{u}_{k+1} = \hat{\boldsymbol{u}}^\star$. Hence with $\|\mathbf{u}_{k+1}\|_{\ell_2} = \|\hat{\boldsymbol{u}}^\star\|_{\ell_2} = 1$ it follows that $\hat{\mathbf{u}}_{k+1} = \mathbf{u}_{k+1} = \hat{\boldsymbol{u}}^\star$, and thus $\hat{\boldsymbol{u}}^\star$ is a fixed point. To establish the stability of the fixed point, define

$$\boldsymbol{G}(\boldsymbol{u}) = \frac{\boldsymbol{g}(\boldsymbol{u})}{\|\boldsymbol{g}(\boldsymbol{u})\|_{\ell_2}}, \quad \boldsymbol{g}(\boldsymbol{u}) = \boldsymbol{u} + \tau[Y_l\, \sigma(Y_l^{\mathrm{T}} \boldsymbol{u}) - \boldsymbol{u} \|Y_l^{\mathrm{T}} \boldsymbol{u}\|_{\ell_1}] \tag{8.19}$$

so that (5.11) is given by

$$\hat{\mathbf{u}}_{k+1} = \boldsymbol{G}(\hat{\mathbf{u}}_k). \tag{8.20}$$

The claimed stability will follows once it is shown that the Jacobian of this mapping evaluated at $\hat{\boldsymbol{u}}^\star$ has spectral radius less than 1 when τ is sufficiently small. The Jacobian is given by

$$\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{u}} = \frac{1}{\|\boldsymbol{g}(\boldsymbol{u})\|_{\ell_2}} \left(I - \frac{\boldsymbol{g}(\boldsymbol{u}) \boldsymbol{g}^{\mathrm{T}}(\boldsymbol{u})}{\|\boldsymbol{g}(\boldsymbol{u})\|_{\ell_2}^2} \right) \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{u}}(\boldsymbol{u}) \tag{8.21}$$

where with $\{j : \hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u} = 0\} = \emptyset$,

$$\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{u}}(\boldsymbol{u}) = (1 - \tau \|Y_l^{\mathrm{T}} \boldsymbol{u}\|_{\ell_1}) I - \tau \boldsymbol{u} \sum_{j=1}^{n} \sigma(\hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \boldsymbol{u}) \hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}}. \tag{8.22}$$

It follows with (8.9) that $\boldsymbol{g}(\hat{\boldsymbol{u}}^\star) = \hat{\boldsymbol{u}}^\star$ and thus $\|\boldsymbol{g}(\hat{\boldsymbol{u}}^\star)\|_{\ell_2} = \|\hat{\boldsymbol{u}}^\star\|_{\ell_2} = 1$. The Jacobian of $\boldsymbol{G}$ at $\hat{\boldsymbol{u}}^\star$ is symmetric and is given by

$$\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{u}}(\hat{\boldsymbol{u}}^\star) = (1 - \tau \|Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star\|)(I - \hat{\boldsymbol{u}}^\star \hat{\boldsymbol{u}}^{\star \mathrm{T}}). \tag{8.23}$$

For $\boldsymbol{v} \in \mathbb{R}^m$ it follows with $0 \leq \|(I - \hat{\boldsymbol{u}}^\star \hat{\boldsymbol{u}}^{\star \mathrm{T}}) \boldsymbol{v}\|_{\ell_2}^2 = \boldsymbol{v}^{\mathrm{T}} (I - \hat{\boldsymbol{u}}^\star \hat{\boldsymbol{u}}^{\star \mathrm{T}}) \boldsymbol{v} \leq \|\boldsymbol{v}\|_{\ell_2}^2$ that the Jacobian satisfies

$$0 \leq \boldsymbol{v}^{\mathrm{T}} \frac{\partial \boldsymbol{G}}{\partial \boldsymbol{u}}(\hat{\boldsymbol{u}}^\star) \boldsymbol{v} \leq (1 - \tau \|Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star\|_{\ell_1}) \|\boldsymbol{v}\|_{\ell_2}^2 < \|\boldsymbol{v}\|_{\ell_2}^2, \quad \forall \boldsymbol{v} \in \mathbb{R}^m \tag{8.24}$$

provided that $1 > \tau \|\hat{\boldsymbol{u}}^{\star \mathrm{T}} Y_l\|_{\ell_1}$. For any such τ the spectral radius of the Jacobian in (8.23) is less than 1. ∎

Remark 3 *Computations demonstrate that the iteration (5.11) converges to a maximizer for G_l even when the condition $\mathcal{S} = \{j : \hat{\boldsymbol{e}}_j^{\mathrm{T}} Y_l^{\mathrm{T}} \hat{\boldsymbol{u}}^\star = 0\} = \emptyset$ is violated.*

9 Application to DCE-MRI Sequences

In this section the proposed methods are applied to the dynamic contrast enhanced magnetic resonance image (DCE-MRI) sequence

http://math.uni-graz.at/keeling/manuskripten/dcemri.mpg

to separate intensity changes due to physiological motion from those due to contrast agent. With such a separation, unavoidable physiological motion may be removed in order to investigate tissues in a stationary state. See also [14] and [16]. To focus on the period in which contrast agent arrives in the imaged tissues, only the first 40 of 134 frames are used for the following decompositions. Each frame consists of an image with 400×400 pixels. Thus, in the notation of Section 2, the data are $Y \in \mathbb{R}^{m \times n}$ with $m = 400^2 \gg 40 = n$. For a static display of the DCE-MRI sequence, representative stages are shown in Fig. 5: exhale and inhale, with and without contrast agent. Specifically, with $\bar{Y}, V, \Lambda$ given by (2.7) and (2.10), the images of Fig. 5 are given by

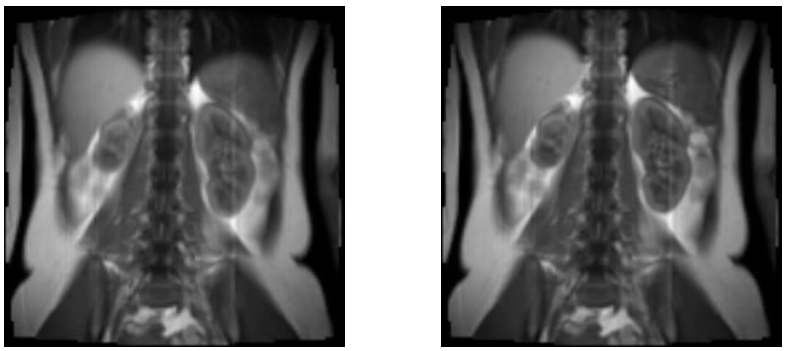
$$\bar{Y} + V \Lambda^{\frac{1}{2}} \boldsymbol{e}_{ij}, \quad i, j = \pm 1, \quad \boldsymbol{e}_{ij} = (i, j, 0, \ldots, 0)^{\mathrm{T}} \tag{9.1}$$

and the images $\bar{Y}$ and

$$\hat{\boldsymbol{v}}_i = V \hat{\boldsymbol{e}}_i, \quad i = 1, 2, \quad (\hat{\boldsymbol{e}}_i)_j = \delta_{ij}, \tag{9.2}$$

are shown in Fig. 6. Brightness changes in relation to the background are seen throughout organs in Fig. 6b, and this suggests that $\hat{\boldsymbol{v}}_1$ represents intensity changes in the DCE-MRI sequence due to contrast agent. On the other hand, brightness changes are seen mainly on the edges of organs in Fig. 6c, and this suggests that $\hat{\boldsymbol{v}}_2$ represents intensity changes in the DCE-MRI sequence due to physiological motion. The image sequence is shown more dynamically in Fig. 7. The graphs in the left column are the time courses $\hat{\boldsymbol{v}}_i^{\mathrm{T}} Y$, $\hat{\boldsymbol{v}}_i^{\mathrm{T}} Y_{\mathrm{s}}$, $\hat{\boldsymbol{v}}_i^{\mathrm{T}} X_{\mathrm{c}}$, $i = 1, 2$, for the raw, sphered and independent data, respectively, where Y_{s} and X_{c} are given by (2.8) and (2.14), respectively. The graphs in the right column are corresponding plots in a phase plane. To determine the most significant independent components, all but the top two principal components were discarded. Then V was replaced by its first two columns, Y_{s} by its first two rows and Λ by a diagonal matrix containing the largest two eigenvalues. Also, the independent images

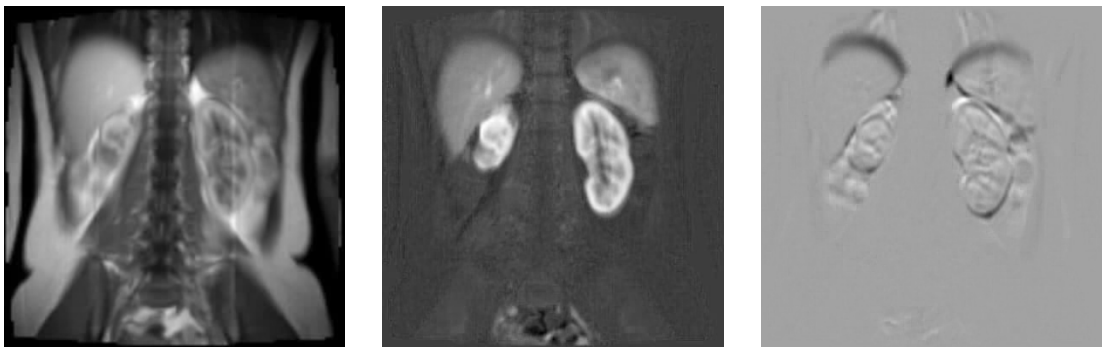
$$V \Lambda^{\frac{1}{2}} U^{\mathrm{T}} Q_i X_{\mathrm{c}}, \quad i = 1, 2, \quad Q_i = \mathrm{diag}\{\hat{\boldsymbol{e}}_i\}, \quad (\hat{\boldsymbol{e}}_i)_j = \delta_{ij}, \tag{9.3}$$



(a) exhale without contrast agent (b) inhale without contrast agent

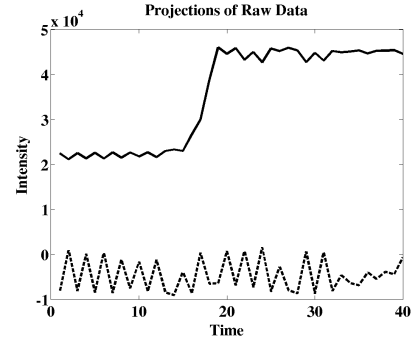
(c) exhale with contrast agent (d) inhale with contrast agent

Figure 5: Representative stages of the DCE-MRI sequence: exhale and inhale with and without contrast agent, where these are defined by (9.1).

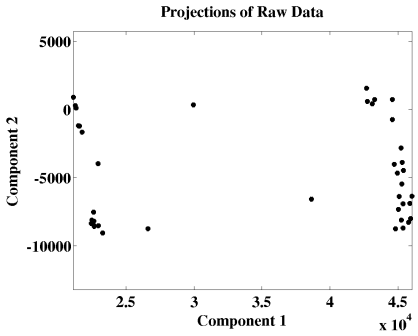


(a) ℓ_2 mean image $\bar{Y}$ (b) ℓ_2 principal component $\hat{\boldsymbol{v}}_1$ (c) ℓ_2 principal component $\hat{\boldsymbol{v}}_2$

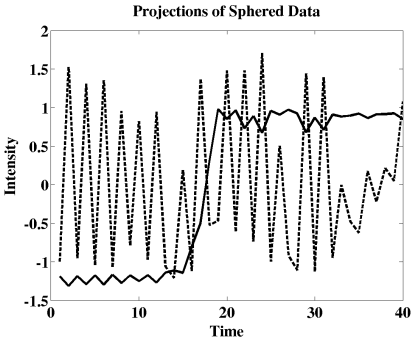
Figure 6: (a) Mean image and (b) - (c) the first two principal components of the DCE-MRI sequence obtained by ℓ_2 methods. Intensity changes in the image sequence associated with contrast agent and with physiological motion are conspicuously apparent in the component $\hat{\boldsymbol{v}}_1$ and $\hat{\boldsymbol{v}}_2$ respectively.



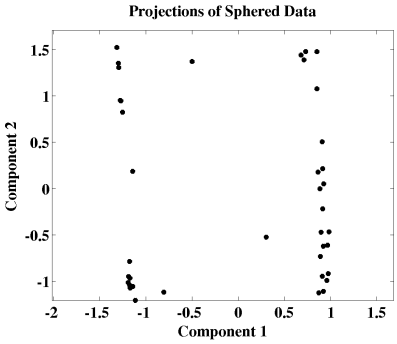
(a) times courses $\hat{\boldsymbol{v}}_i^{\mathrm{T}} Y$, $i = 1, 2$, of raw data



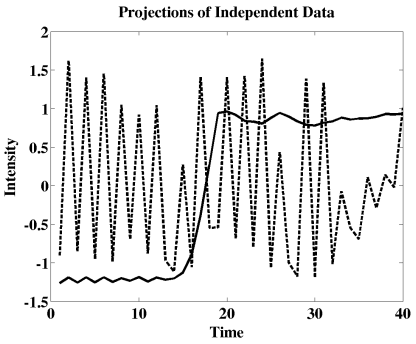
(b) raw data $\hat{\boldsymbol{v}}_i^{\mathrm{T}} Y$, $i = 1, 2$, in phase plane



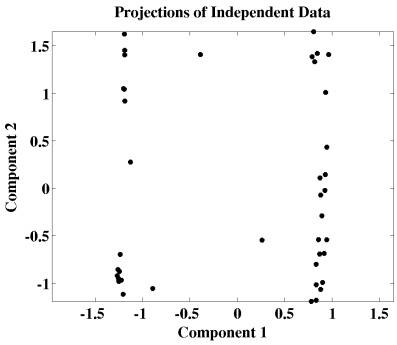
(c) times courses $\hat{\boldsymbol{v}}_i^{\mathrm{T}} Y_{\mathrm{s}}$, $i = 1, 2$, of sphered data



(d) sphered data $\hat{\boldsymbol{v}}_i^{\mathrm{T}} Y_{\mathrm{s}}$, $i = 1, 2$, in phase plane

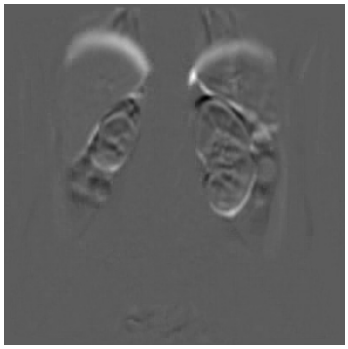


(e) times courses $\hat{\boldsymbol{v}}_i^{\mathrm{T}} X_{\mathrm{c}}$, $i = 1, 2$, of independent data



(f) independent data $\hat{\boldsymbol{v}}_i^{\mathrm{T}} X_{\mathrm{c}}$, $i = 1, 2$, in phase plane

(g) independent component 1



(h) independent component 2

Figure 7: Representation of components of raw, sphered and independent data for the DCE-MRI sequence. These have been determined by ℓ_2 based methods.

are shown in Figs. 7g and 7h. As explained in connection with Fig. 6, these can be associated respectively with intensity changes in the DCE-MRI sequence due to contrast agent and to physiological motion. Note that there are only small differences between Figs. 7c and 7e, between Figs. 7d and 7f, between Figs. 6b and 7g and between Figs. 6c and 7h. Thus, the separation of intensity changes due to physiological motion from those due to contrast agent is achieved here already with the sphered data. Hence the transformation to independent data had little effect for this particular example. Recall from Section 2 that the order and the sign of ICA components are not uniquely determined.

This separation will now be considered in the presence of outliers. As seen in the full DCE-MRI sequence, an excessivley bright frame may appear suddenly. To simulate this effect, intensities of the final frame of the sequence are increased by a constant factor. Then the same methods used for Fig. 7 are applied to the corrupted data, and the results are shown with the same format as used for Fig. 8. The corrupted data may be seen at the final time shown in the graphs of the first column of Fig. 8. Also the outlier is conspicuous in the phase plane graphs in the right column of Fig. 8. Finally, the images defined by (9.3) with the corrupted data are shown in Figs. 8g and Fig. 8h. Since these clearly differ from their counterparts in Figs. 7g and Fig. 7h, the presence of the single outlier has corrupted the separation of intensity changes due to physiological motion from those due to contrast agent.

For comparison, the ℓ_1 based methods of Sections 3 – 5 are now applied to the corrupted data, and the results are shown with the same format as used for Fig. 9. Now the matrices $\bar{V}$, Y_c, V, Λ, Y_s, and U are understood as explained in Sections 3 – 5 as well as in Remark 2. As in the previous cases, all but the top two principal components are discarded and the respective matrices are reduced correspondingly to have only two rows or columns or both. Then (9.2) and (9.3) apply with these ℓ_1 based matrices. As before, the corrupted data may be seen at the final time shown in the graphs of the first column of Fig. 9. Also the outlier is conspicuous in the phase plane graphs in the right column of Fig. 9. Finally, the images defined by (9.3) with the corrupted data are shown in Figs. 9g and Fig. 9h. Note the similarities between Figs. 9e – 9h and their counterparts in Figs. 7e – 7h. On this basis, the ℓ_1 methods can be seen to have successfully separated intensity changes due to physiological motion from those due to contrast agent in spite of the outlier. With this separation, the data are projected onto the single independent component of Fig. 9g using (2.21) to produce the following transformed DCE-MRI sequence manifesting contrast changes free of physiological motion,

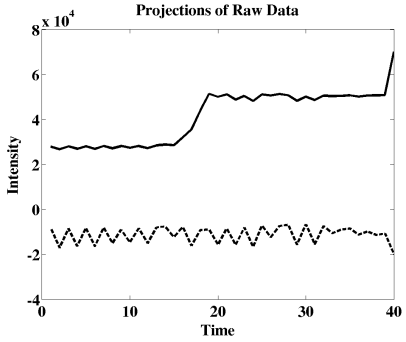
http://math.uni-graz.at/keeling/manuskripten/dcemri_ica.mpg.

10 Conclusion

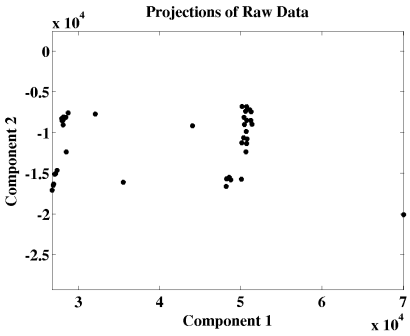
In this work robust measures have been introduced for centering complex data in the presence of outliers and for determining principal and independent components of such data. The approach to centering is to use the geometric median. The approach for determining principal components is to find best fit lines through the data, where each line minimizes the sum of distances (not squared) to data points in the subspace orthogonal to other components. The approach for determining independent components is first to sphere the data so that the corresponding axes are aligned with clusters, and then to determine independent axes as those which separate sphered clusters as much as possible. This separation is accomplished by maximizing an ℓ_1 counterpart to Rayleigh quotients. To optimize the respective merit functions, iterative primal-dual based methods were proposed and their convergence was proved. Illustrative examples were presented to demonstrate the benefits of the robust approaches. Finally, the proposed methods were applied to a DCE-MRI sequence to separate intensity changes due to physiological motion from those due to contrast agent, and benefits of the robust methods have been demonstrated with respect to this realistic example.

11 Acknowledgement

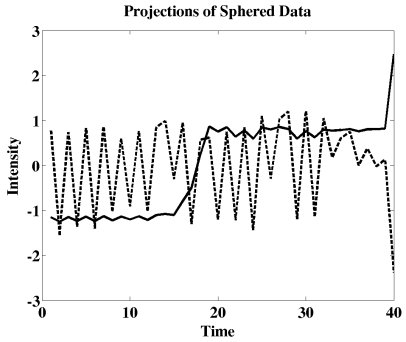
The authors wish to express their appreciation for many fruitful discussions with Gernot Reishofer of the Radiology Clinic of the Graz Medical University, who also provided data for this work. Also,



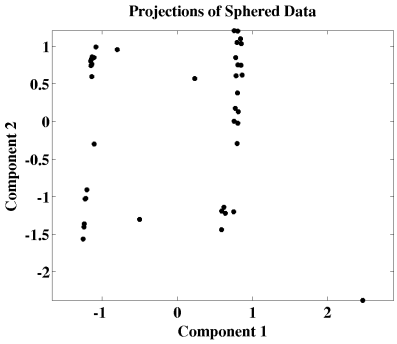
(a) times courses $\hat{\boldsymbol{v}}_i^{\mathrm{T}} Y$, $i = 1, 2$, of raw data



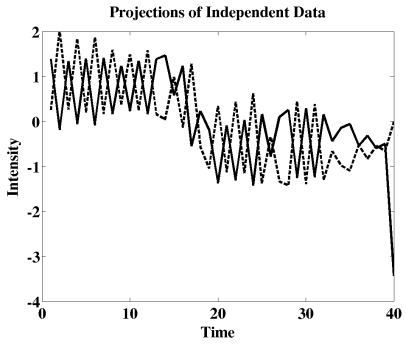
(b) raw data $\hat{\boldsymbol{v}}_i^{\mathrm{T}} Y$, $i = 1, 2$, in phase plane



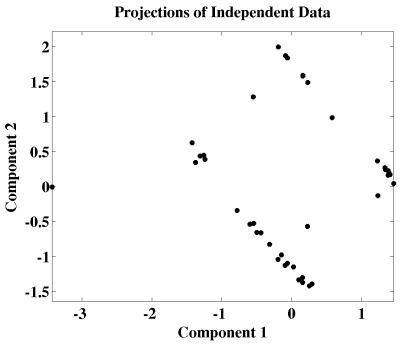
(c) times courses $\hat{\boldsymbol{v}}_i^{\mathrm{T}} Y_{\mathrm{s}}$, $i = 1, 2$, of sphered data



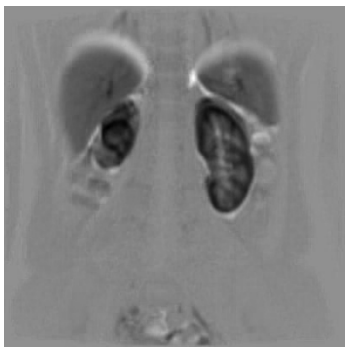
(d) sphered data $\hat{\boldsymbol{v}}_i^{\mathrm{T}} Y_{\mathrm{s}}$, $i = 1, 2$, in phase plane



(e) times courses $\hat{\boldsymbol{v}}_i^{\mathrm{T}} X_{\mathrm{c}}$, $i = 1, 2$, of independent data



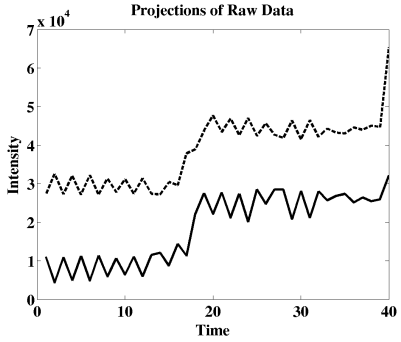
(f) independent data $\hat{\boldsymbol{v}}_i^{\mathrm{T}} X_{\mathrm{c}}$, $i = 1, 2$, in phase plane



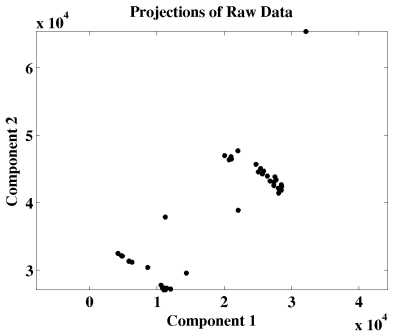
(g) independent component 1

(h) independent component 2

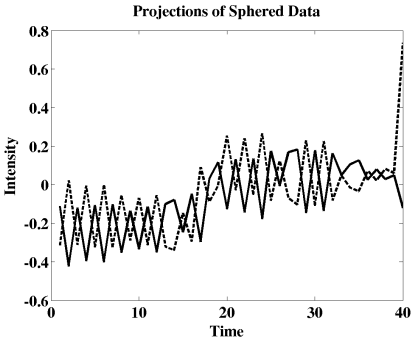
Figure 8: Representation of components of raw, sphered and independent data for the DCE-MRI sequence with a single outlier introduced at the final time. These have been determined by ℓ_2 based methods.



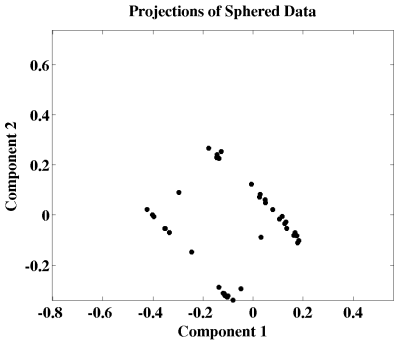
(a) times courses $\hat{\boldsymbol{v}}_i^{\mathrm{T}} Y$, $i = 1, 2$, of raw data



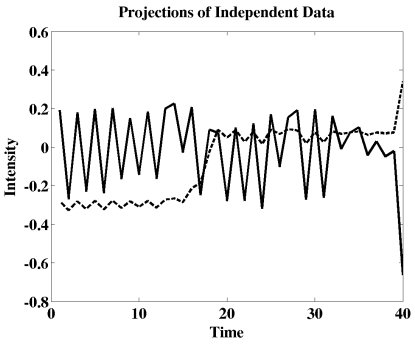
(b) raw data $\hat{\boldsymbol{v}}_i^{\mathrm{T}} Y$, $i = 1, 2$, in phase plane



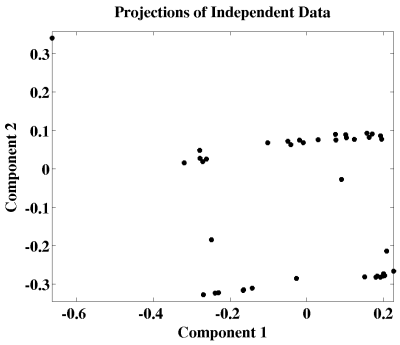
(c) times courses $\hat{\boldsymbol{v}}_i^{\mathrm{T}} Y_{\mathrm{s}}$, $i = 1, 2$, of sphered data



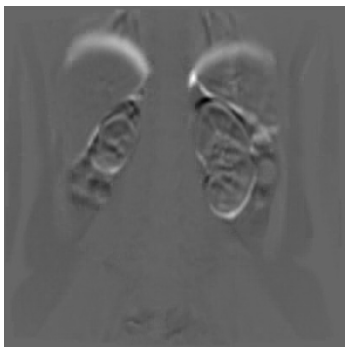
(d) sphered data $\hat{\boldsymbol{v}}_i^{\mathrm{T}} Y_{\mathrm{s}}$, $i = 1, 2$, in phase plane



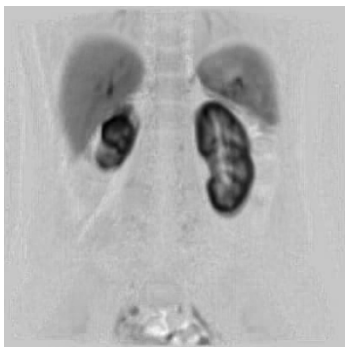
(e) times courses $\hat{\boldsymbol{v}}_i^{\mathrm{T}} X_{\mathrm{c}}$, $i = 1, 2$, of independent data



(f) independent data $\hat{\boldsymbol{v}}_i^{\mathrm{T}} X_{\mathrm{c}}$, $i = 1, 2$, in phase plane



(g) independent component 1



(h) independent component 2

Figure 9: Representation of components of raw, sphered and independent data for the DCE-MRI sequence with a single outlier introduced at the final time. These have been determined by ℓ_1 based methods.

the authors gratefully acknowledge support from the Austrian Science Fund *Fonds zur Förderung der Wissenschaftlichen Forschung* (FWF) under grant SFB F032.

References

[1] R.B. ASH, *Real Analysis and Probability*, Academic Press, New York, 1972.

[2] H.H. BAUSCHKE and P.L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, 2010.

[3] J.P. BROOKS, J.H. DULÁ and E.L. BOONE, *A pure L_1-norm principal comoponent analysis*, Computational Statistics and Data Analysis, Vol. 61, pp. 83 – 98, May 2013.

[4] P. BOSE, A. MAHESHWARI and P. MORIN, *Fast approximations for sums of distances, clustering and the FermatWeber problem*, Computational Geometry: Theory and Applications, Vol. 24, No. 3, pp. 135 – 146, 2003.

[5] T.T. CAI, Z. MA and Y. WU, *Sparse PCA: Optimal rates and adaptive estimation*, Annals of Statistics, Vol. 41, No. 6, pp. 3074 – 3110, 2013.

[6] A. CHAMBOLLE and T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging*, JMIV, Vol. 40, No. 1, pp. 120 – 145, 2011.

[7] C. DING and D. ZHOU and X. HE and H. ZHA, *R_1-pca: Rotation invariant L_1-norm principal component analysis for robust subspace factorization*, Proceedings of the 23rd International Coference on Machine Learning, pp. 281 – 288, 2006.

[8] Y. DODGE and V. ROUSSON, *Multivariate L_1 Mean*, Metrika, Vol. 49, No. 2, pp. 127 – 134, 1999.

[9] I. EKELAND and R. TÉMAM, *Convex Analysis and Variational Problems*, North-Holland, Amsterdam-Oxford, 1976.

[10] A. HYVÄRINEN, J. KARHUNEN and E. OJA, *Independent Component Analysis*, New York, Wiley, 2001.

[11] I.T. JOLLIFFE, *Principal Component Analysis*, New York, Springer, 1986.

[12] N. KWAK, *Principal component analysis based on $L1$-norm maximization*, IEEE Transactions on Pattern Analysis and Machine Intelligence Vol. 30, pp. 1672 – 1680, 2008.

[13] E. WEISZFELD and F. PLASTRIA, *On the point for which the sum of the distances to n given points is minimum*, Ann. Oper. Res., Vol. 167, pp. 7 – 41, 2009.

[14] G. REISHOFER, F. FAZEKAS, S.L. KEELING, C. ENZINGER, F. PAYER, J. SIMBRUNNER and R. STOLLBERGER, *Minimizing Macrovessel Signal in Cerebral Perfusion Imaging using Independent Component Analysis*, Magnetic Resonance in Medicine, Vol. 57, No. 2, pp. 278 – 288, 2007.

[15] E. WEISZFELD, *Sur le point pour lequel la somme des distances de n points donnés est minimum*, Tôhoku Mathematical Journal, Vol. 43, pp. 355 – 386, 1937.

[16] G. WOLLNY, P. KELLMAN, A. SANTOS and M.J. LEDESMA-CARBAYO, *Automatic motion compensation of free breathing acquired myocardial perfusion data by using independent component analysis*, Medical Image Analysis, Vol. 16, pp. 1015 – 1028, 2012.

[17] V. ZARZOSO, *Robust Independent Component Analysis by Iterative Maximization of the Kurtosis Contrast with Algebraic Optimal Step Size*, IEEE Trans. on Neural Networks, Vol. 21, No. 2, pp. 248 – 261, Feb. 2010.

A Alternative Computation of the Geometric Median

The purpose of this section is to show that the Chambolle-Pock Algorithm [6] can be adapted for computing the geometric median. This approach is subsequently compared with that of the primal-dual algorithm of Section 3. Using elements of convex analysis [9], problem (3.4) may be reformulated as

$$\min_{\vec{\boldsymbol{\mu}}\in\mathbb{R}^{m\times n}} F(\vec{\boldsymbol{\mu}}) + G(\vec{\boldsymbol{\mu}}) \tag{A.1}$$

where

$$F(\vec{\boldsymbol{\mu}}) = \sum_{j=1}^{n} f_j(\boldsymbol{\mu}_j), \quad f_j(\boldsymbol{\mu}_j) = \|\boldsymbol{\mu}_j - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}, \quad \vec{\boldsymbol{\mu}} = \{\boldsymbol{\mu}_j\}_{j=1}^{n}, \quad \boldsymbol{\mu}_j \in \mathbb{R}^m \tag{A.2}$$

and with the indicator function

$$I_{\mathcal{C}}(\vec{\boldsymbol{\mu}}) = \begin{cases} 0, & \vec{\boldsymbol{\mu}} \in \mathcal{C} \\ \infty, & \text{otherwise} \end{cases} \tag{A.3}$$

$$G(\vec{\boldsymbol{\mu}}) = I_{\mathcal{C}}(\vec{\boldsymbol{\mu}}), \quad \mathcal{C} = \{\vec{\boldsymbol{\mu}} = \{\boldsymbol{\mu}_j\}_{j=1}^{n} : \boldsymbol{\mu}_{j_1} = \boldsymbol{\mu}_{j_2}, 1 \le j_1, j_2 \le n\}. \tag{A.4}$$

The convex conjugate f_j^* is given by

$$\begin{aligned} f_j^*(\boldsymbol{\delta}_j) &= \sup_{\boldsymbol{\mu}\in\mathbb{R}^m} \left[\boldsymbol{\delta}_j^{\mathrm{T}}\boldsymbol{\mu} - \|\boldsymbol{\mu} - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}\right] = \sup_{\boldsymbol{\mu}\in\mathbb{R}^m} \left[\boldsymbol{\delta}_j^{\mathrm{T}}(\boldsymbol{\mu} - Y\hat{\boldsymbol{e}}_j) - \|\boldsymbol{\mu} - Y\hat{\boldsymbol{e}}_j\|_{\ell_2}\right] + \boldsymbol{\delta}_j^{\mathrm{T}} Y\hat{\boldsymbol{e}}_j \\ &= I_B(\boldsymbol{\delta}_j) + \boldsymbol{\delta}_j^{\mathrm{T}} Y\hat{\boldsymbol{e}}_j \end{aligned} \tag{A.5}$$

where

$$B = \{\boldsymbol{\delta} : \|\delta\|_{\ell_2} \le 1\}, \quad I_B(\boldsymbol{\delta}) = \begin{cases} 0, & \boldsymbol{\delta} \in B \\ \infty, & \text{otherwise.} \end{cases} \tag{A.6}$$

The convex conjugate F^* is given componentwise according to

$$F^*(\vec{\boldsymbol{\delta}}) = \sum_{j=1}^{n} I_B(\boldsymbol{\delta}_j) + \boldsymbol{\delta}_j^{\mathrm{T}} Y\hat{\boldsymbol{e}}_j, \quad \vec{\boldsymbol{\delta}} = \{\boldsymbol{\delta}_j\}_{j=1}^{n} \tag{A.7}$$

The primal problem (A.1) is reformulated as a saddle point problem according to

$$\min_{\vec{\boldsymbol{\mu}}\in\mathbb{R}^{m\times n}} \left[\max_{\vec{\boldsymbol{\delta}}\in\mathbb{R}^{m\times n}} \vec{\boldsymbol{\delta}} : \vec{\boldsymbol{\mu}} - F^*(\vec{\boldsymbol{\delta}})\right] + G(\vec{\boldsymbol{\mu}}) = \min_{\vec{\boldsymbol{\mu}}\in\mathbb{R}^{m\times n}} \max_{\vec{\boldsymbol{\delta}}\in\mathbb{R}^{m\times n}} \left[\sum_{j=1}^{n} \boldsymbol{\delta}_j^{\mathrm{T}}(\boldsymbol{\mu}_j - Y\hat{\boldsymbol{e}}_j) - I_B(\boldsymbol{\delta}_j) + I_{\mathcal{C}}(\vec{\boldsymbol{\mu}})\right] \tag{A.8}$$

For the Chambolle-Pock Algorithm [6]

$$\begin{cases} \vec{\boldsymbol{\delta}}^{k+1} &= (I + \varsigma\partial F^*)^{-1}(\vec{\boldsymbol{\delta}}^{k} + \varsigma\vec{\boldsymbol{\nu}}^{k}) \\ \vec{\boldsymbol{\mu}}^{k+1} &= (I + \varsigma\partial G)^{-1}(\vec{\boldsymbol{\mu}}^{k} - \tau\vec{\boldsymbol{\delta}}^{k+1}) \\ \vec{\boldsymbol{\nu}}^{k+1} &= \vec{\boldsymbol{\mu}}^{k+1} + \theta(\vec{\boldsymbol{\mu}}^{k+1} - \vec{\boldsymbol{\mu}}^{k}) \end{cases} \qquad 0 < \tau\varsigma < 1, \quad \theta \in [0,1] \tag{A.9}$$

the operator $(I + \varsigma\partial G)^{-1}$ is characterized by

$$\vec{\boldsymbol{y}}(I + \varsigma\partial G)^{-1}\vec{\boldsymbol{z}} = \mathcal{P}_{\mathcal{C}}\vec{\boldsymbol{z}} \quad \text{or} \quad \boldsymbol{y}_j = \boldsymbol{y} = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{z}_j, \quad \vec{\boldsymbol{y}} = \{\boldsymbol{y}_j\}_{j=1}^{n}, \quad \vec{\boldsymbol{z}} = \{\boldsymbol{z}_j\}_{j=1}^{n} \tag{A.10}$$

and the operator $(I + \varsigma\partial F^*)^{-1}$ is characterized componentwise as

$$\vec{\boldsymbol{y}} \in (I + \varsigma\partial F^*)\vec{\boldsymbol{z}}, \quad \Leftrightarrow \quad \boldsymbol{y}_j \in (I + \varsigma\partial f_j^*)\boldsymbol{z}_j, \quad \vec{\boldsymbol{y}} = \{\boldsymbol{y}_j\}_{j=1}^{n}, \quad \vec{\boldsymbol{z}} = \{\boldsymbol{z}_j\}_{j=1}^{n}. \tag{A.11}$$

Hence

$$\frac{\boldsymbol{y}_j - \boldsymbol{z}_j}{\varsigma} \in \partial[I_B(\boldsymbol{z}_j) + \boldsymbol{z}_j^{\mathrm{T}} Y\hat{\boldsymbol{e}}_j] \tag{A.12}$$

or equivalently

$$\begin{aligned} \left\| \frac{\boldsymbol{y}_j - \boldsymbol{z}_j}{\varsigma} - Y\hat{\boldsymbol{e}}_j \right\|_{\ell_2} + I_B(\boldsymbol{z}_j) + \boldsymbol{z}_j^{\mathrm{T}} Y\hat{\boldsymbol{e}}_j &= \\ f_j\left(\frac{\boldsymbol{y}_j - \boldsymbol{z}_j}{\varsigma}\right) + f_j^*(\boldsymbol{z}_j) &= \left(\frac{\boldsymbol{y}_j - \boldsymbol{z}_j}{\varsigma}\right)^{\mathrm{T}} \boldsymbol{z}_j \end{aligned} \tag{A.13}$$

i.e.,

$$\|(\boldsymbol{y}_j - \varsigma Y\hat{\boldsymbol{e}}_j) - \boldsymbol{z}_j\|_{\ell_2} + I_B(\boldsymbol{z}_j) = [(\boldsymbol{y}_j - \varsigma Y\hat{\boldsymbol{e}}_j) - \boldsymbol{z}_j]^{\mathrm{T}} \boldsymbol{z}_j \tag{A.14}$$

The case that $\|\boldsymbol{y}_j - \varsigma Y\hat{\boldsymbol{e}}_j\|_{\ell_2} = 0$ holds gives $\boldsymbol{z}_j = 0$. On the other hand, setting $\boldsymbol{z}_j = \eta(\boldsymbol{y}_j - \varsigma Y\hat{\boldsymbol{e}}_j)$ for $|\eta| \leq 1/\|\boldsymbol{y}_j - \varsigma Y\hat{\boldsymbol{e}}_j\|_{\ell_2}$ gives

$$|1-\eta|\|\boldsymbol{y}_j - \varsigma Y\hat{\boldsymbol{e}}_j\|_{\ell_2} = (1-\eta)\eta\|\boldsymbol{y}_j - \varsigma Y\hat{\boldsymbol{e}}_j\|_{\ell_2}^2 \tag{A.15}$$

which is solved by

$$\eta = \begin{cases} 1/\|\boldsymbol{y}_j - \varsigma Y\hat{\boldsymbol{e}}_j\|_{\ell_2}, & \|\boldsymbol{y}_j - \varsigma Y\hat{\boldsymbol{e}}_j\|_{\ell_2} \geq 1 \\ 1, & \|\boldsymbol{y}_j - \varsigma Y\hat{\boldsymbol{e}}_j\|_{\ell_2} < 1 \end{cases} \tag{A.16}$$

giving

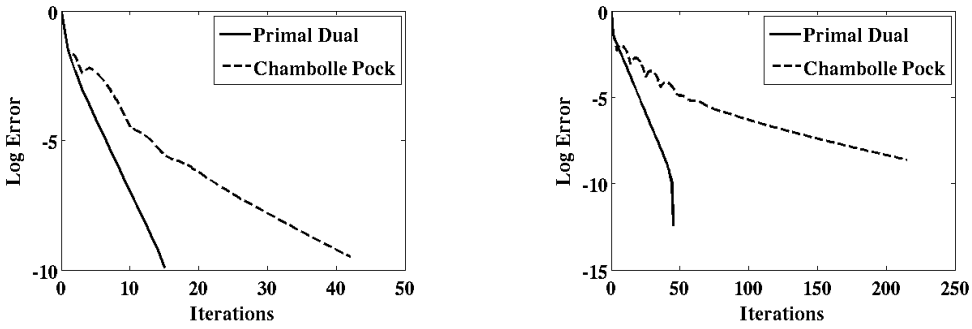
$$\boldsymbol{z}_j = \begin{cases} \dfrac{\boldsymbol{y}_j - \varsigma Y\hat{\boldsymbol{e}}_j}{\|\boldsymbol{y}_j - \varsigma Y\hat{\boldsymbol{e}}_j\|_{\ell_2}}, & \|\boldsymbol{y}_j - \varsigma Y\hat{\boldsymbol{e}}_j\|_{\ell_2} \geq 1 \\ \boldsymbol{y}_j - \varsigma Y\hat{\boldsymbol{e}}_j, & \|\boldsymbol{y}_j - \varsigma Y\hat{\boldsymbol{e}}_j\|_{\ell_2} < 1 \end{cases} = P_B(\boldsymbol{y}_j - \varsigma Y\hat{\boldsymbol{e}}_j) \tag{A.17}$$

Finally, the Chambolle-Pock Algorithm can be written as $\boldsymbol{\mu}_l, \boldsymbol{\nu}_l \in \mathbb{R}^m$, $D_l \in \mathbb{R}^{m\times n}$

$$\begin{cases} D_{l+1}\hat{\boldsymbol{e}}_j &= P_B(D_l\hat{\boldsymbol{e}}_j + \varsigma(\boldsymbol{\nu}_l - Y\hat{\boldsymbol{e}}_j)), \quad 1 \leq j \leq n \\ \boldsymbol{\mu}_{l+1} &= \boldsymbol{\mu}_l - \tau/n \sum_{j=1}^n D_{l+1}\hat{\boldsymbol{e}}_j \\ \boldsymbol{\nu}_{l+1} &= \boldsymbol{\mu}_{l+1} + \theta(\boldsymbol{\mu}_{l+1} - \boldsymbol{\mu}_l) \end{cases} \qquad P_B(\boldsymbol{\nu}) = \begin{cases} \boldsymbol{\nu}/\|\boldsymbol{\nu}\|_{\ell_2}, & \|\boldsymbol{\nu}\|_{\ell_2} > 1 \\ \boldsymbol{\nu}, & \|\boldsymbol{\nu}\|_{\ell_2} \leq 1 \end{cases} \tag{A.18}$$

The parameters satisfy $\theta = 1$ and $0 < \varsigma, \tau < 1$ with $\tau \gg 1$ to accelerate convergence of $\boldsymbol{\mu}_l$ to the geometric median. This faster convergence gives the one-dimensional median in case the data are colinear as formulated in (6.1).

As can be seen in Fig. 10, the primal-dual scheme (3.6) – (3.7) has been found to converge faster than



(a) convergence for temporal geometric median

(b) convergence for spatial geometric median

Figure 10: Convergence histories for the computation of the temporal and spatial geometric medians of the DCE-MRI data of Section 9 according to the primal-dual scheme (3.6) – (3.7) and the Chambolle-Pock scheme (A.18). In both (a) and (b) convergence of the primal-dual scheme is shown with solid curves while the convergence of the Chambolle-Pock scheme is shown with dashed curves. In (a) the geometric median is computed with respect to time, where components of the result correspond to spatial intensities. In (b) the geometric median is computed with respect to space, where components of the result correspond to temporal intensities.

the Chambolle-Pock scheme (A.18). Shown in Fig. 10 are convergence histories for the two schemes

during their computation of the temporal and spatial geometric medians of the full data set cited in Section 9. In both graphs of Fig. 10 convergence of the primal-dual scheme is shown with solid curves while the convergence of the Chambolle-Pock scheme is shown with dashed curves. Numerical parameters have been chosen to provide the most rapid convergence obtainable with the respective schemes. In Fig. 10a the geometric median is computed over 134 times points, and the 400^2 components of the result correspond to spatial intensities. In Fig. 10b the geometric median is computed over 400^2 spatial points, and the 134 components of the result correspond to temporal intensities. Note that the spatial geometric median is not used in Section 9, but it is addressed here for comparison purposes since some authors tend to reverse the role of space and time as discussed at the end of Section 2. The presentation of the spatial geometric median also highlights the general trend that the differences in speed between the two schemes increases rapidly with the number of points over which the geometric median is computed. On the basis of the results shown in Fig. 10, the primal-dual scheme (3.6) – (3.7) has been proposed in this work instead of the Chambolle-Pock scheme (A.18).