

# A NOTE ON THE LOCALLY LINEAR EMBEDDING ALGORITHM

WOJCIECH CHOJNACKI<sup>1,2,\*</sup> and MICHAEL J. BROOKS<sup>1,†</sup>

<sup>1</sup>*School of Computer Science*

*The University of Adelaide, Adelaide, SA 5005, Australia*

<sup>2</sup>*Wydział Matematyczno-Przyrodniczy, Szkoła Nauk Ścisłych*

*Uniwersytet Kardynała Stefana Wyszyńskiego, Dewajtis 5, 01-815 Warszawa, Poland*

*\*wojciech.chojnacki@adelaide.edu.au*

*†michael.brooks@adelaide.edu.au*

The paper presents mathematical underpinnings of the locally linear embedding technique for data dimensionality reduction. It is shown that a cogent framework for describing the method is that of optimisation on a Grassmann manifold. The solution delivered by the algorithm is characterised as a constrained minimiser for a problem in which the cost function and all the constraints are defined on such a manifold. The role of the internal gauge symmetry in solving the underlying optimisation problem is illuminated.

*Keywords:* Dimensionality reduction; locally linear embedding; Stiefel manifold; Grassmann manifold; optimisation; gauge freedom; gauge fixing

## 1. Introduction

High-dimensional data can often be converted to low-dimensional data with little or no fundamental loss of information. A simple and widely used method for dimensionality reduction is principal component analysis (PCA). The method represents data points by their respective orthogonal projections on a subspace of low dimension spanned by the directions (also called components, features, factors or sources) of greatest variance in the data set. The locally linear embedding (LLE)<sup>12</sup> and ISOMAP<sup>16</sup> algorithms are two recently proposed non-linear generalisations of PCA. Originally developed for visualisation purposes, these two methods project (embed) high-dimensional data into a two- or low-dimensional subspace by extracting meaningful components in a non-linear fashion. For a brief but informative overview of a class of dimensionality reduction methods that includes PCA, LLE and ISOMAP, see Ref. 14.

This paper focuses on the all-important LLE scheme, the utility of which can hardly be overestimated. Alongside earlier applications in visualisation<sup>4,12,13</sup> and classification,<sup>17</sup> the scheme has most recently found use to such tasks as 3D-object pose estimation,<sup>20</sup> face membership authentication,<sup>11</sup> multipose face synthesis,<sup>18</sup> facial animation,<sup>10</sup> image denoising,<sup>15</sup> hyperspectral image processing,<sup>8</sup> digital watermarking,<sup>6</sup> feature extraction,<sup>7</sup> gait recognition,<sup>9</sup> and manifold learning<sup>19</sup> — to name a few. In most applications, LLE is invoked as a ready-to-use dimensionality

reduction tool. In contrast, this article touches upon some theoretical issues regarding the foundational basis of the algorithm.

The paper gives a detailed account of the mathematical underpinnings of LLE. It is shown that LLE solves a constrained optimisation problem that is usefully viewed as a problem posed on a certain non-Euclidean space, namely, a Grassmann manifold. Such a manifold is the collection of all linear subspaces of fixed dimension in some ambient linear space. Optimisation on manifolds related to the orthogonal group such as Stiefel and Grassmann manifolds frequently appear in the context of neural networks, signal processing, pattern recognition, and computer vision.<sup>5</sup> The benefit of the Grassmann manifold-based formulation of LLE's optimisation problem is in allowing one to see more clearly an internal symmetry, or gauge freedom, related to the arbitrariness of the choice of coordinates in which to naturally state the problem. This paper demonstrates that once recognised, the gauge freedom can be used advantageously to isolate a solution of the optimisation problem. The present exposition of LLE can be seen as an expansion of the presentations given in Refs. 2 and 13.

The paper is arranged as follows. The next section sketches the main ideas behind the working of LLE. The following section presents auxiliary material needed to formulate LLE's optimisation problem. The subsequent section describes the structure of the optimisation problem and a path to its solution. It is here that the main contribution of this note resides, this being an illumination of the role played by the gauge degrees of freedom in isolating the solution. Following concluding remarks, two appendices provide some matrix calculations, which — with a view to self-reliance — are performed in a detailed, explicit manner.

## 2. Main Ideas

The locally linear embedding algorithm assumes that a high-dimensional data set lies on, or near to, a smooth low-dimensional manifold. Small patches of the manifold, each containing a fraction of the data set, can be equipped with individual local co-ordinates. The high-dimensional co-ordinates of each patch can be mapped into corresponding local co-ordinates by means of an essentially linear transformation. LLE attempts to find a global transformation of the high-dimensional co-ordinates into low-dimensional ones by exploiting adjacency information about closely located data points, this information being a form of summarisation of the local transformations between the high- and low-dimensional coordinates.

Suppose that the data set comprises vectors  $\mathbf{x}_1, \dots, \mathbf{x}_N \in \mathbb{R}^D$ . In the first step, for each  $1 \leq i \leq N$ , nearest neighbours of  $\mathbf{x}_i$  are identified by using a preselected criterion for close proximity and further indexed by a set  $\mathcal{N}(i) \subset \{1, \dots, N\}$ . In the second step, weights  $\{w_{ij}\}_{j \in \mathcal{N}(i)}$  are found that optimally reconstruct  $\mathbf{x}_i$  from

its nearest neighbours. These weights minimise the local reconstruction error

$$E_{\text{loc}}^{(i)}(\{w_{ij}\}_{j \in \mathcal{N}(i)}) = \left\| \mathbf{x}_i - \sum_{j \in \mathcal{N}(i)} w_{ij} \mathbf{x}_j \right\|^2,$$

with  $\|\cdot\|$  the Euclidean norm, subject to the condition

$$\sum_{j \in \mathcal{N}(i)} w_{ij} = 1. \quad (1)$$

A key property of the optimal weights is that they are invariant to three types of transformation:

- (1) **Scaling.** Multiplying all co-ordinates by a scalar factor scales the errors  $E_{\text{loc}}^{(i)}$  uniformly and hence yields the same weights.
- (2) **Orthogonal transformation.** Distances are invariant to rotation and mirror reflection and so too is each  $E_{\text{loc}}^{(i)}$ .
- (3) **Translation.** The weights are constrained to sum to one, so an offset to all co-ordinates does not affect the value of any  $E_{\text{loc}}^{(i)}$ .

Suppose that the data points are sampled densely from the underlying low-dimensional manifold. Then, for each point  $\mathbf{x}_i$ , there exists a linear map composed of a translation, rotation and scaling, that maps the high-dimensional co-ordinates of a close neighbourhood of  $\mathbf{x}_i$  to corresponding local co-ordinates on the manifold. Since the weights computed in the high-dimensional space are invariant to the three constituent mappings, it is natural to take these weights as a basis for the reconstruction of the local co-ordinates. In fact, *all* local neighbourhoods can be reconstructed simultaneously if a specific optimisation problem is solved. The cost function for this problem measures how well low-dimensional co-ordinates of any given point  $\mathbf{y}_i \in \mathbb{R}^d$  are reconstructed from the neighbouring points  $\{\mathbf{y}_j\}_{j \in \mathcal{N}(i)}$  using the weights computed in the previous step; here  $d$  is a dimension index fixed beforehand, usually at a value much smaller than  $D$ . More specifically, in the third step, LLE minimises the reconstruction error

$$E_{\text{lle}}(\mathbf{y}_1, \dots, \mathbf{y}_N) = \sum_{i=1}^N \left\| \mathbf{y}_i - \sum_{j \in \mathcal{N}(i)} w_{ij} \mathbf{y}_j \right\|^2.$$

This optimisation is similar to that in the first step, except that now the weights are fixed and the low-dimensional co-ordinates are sought. To obtain an essentially unique solution, the  $\mathbf{y}_i$  are constrained to have zero mean and an identity covariance matrix.

### 3. Computing the Adjacency Matrix

It is convenient to put all candidate weights into a single  $N \times N$  *adjacency matrix*  $\mathbf{W} = [w_{ij}]_{1 \leq i, j \leq N}$ . For each  $1 \leq i \leq N$ , let  $\mathbf{w}_i^{\text{T}} = [w_{i1}, \dots, w_{iN}]$  denote the  $i$ th row of  $\mathbf{W}$ . Its transpose,  $\mathbf{w}_i$ , is by definition a sparse column vector whose essential

entries are labelled by the members of the indexing set  $\mathcal{N}(i)$ . More precisely, if  $N(i)$  is the cardinality of  $\mathcal{N}(i)$  and  $\{n_1, \dots, n_{N(i)}\}$  is the increasing enumeration of  $\mathcal{N}(i)$ , then  $\mathbf{w}_i$  is determined by the sub-vector

$$\tilde{\mathbf{w}}_i = [w_{in_1}, \dots, w_{in_{N(i)}}]^\top,$$

with all the entries of  $\mathbf{w}_i$  not entering  $\tilde{\mathbf{w}}_i$  being set equal to zero; this latter property can of course be expressed by the condition

$$w_{ij} = 0 \quad \text{for } j \notin \mathcal{N}(i). \quad (2)$$

It is obvious that the optimal  $\mathbf{W}$  is the composite of the individually computed optimal rows. Determining the optimal  $\mathbf{w}_i$ , for each  $1 \leq i \leq N$ , is equivalent to determining the optimal  $\tilde{\mathbf{w}}_i$ .

Given  $1 \leq i \leq N$ , define a  $N \times N(i)$  matrix  $\tilde{\mathbf{X}}_i$  by

$$\tilde{\mathbf{X}}_i = [\mathbf{x}_{n_1} - \mathbf{x}_i, \dots, \mathbf{x}_{n_{N(i)}} - \mathbf{x}_i].$$

In view of (1),

$$\begin{aligned} \tilde{\mathbf{X}}_i \tilde{\mathbf{w}}_i &= \sum_{k=1}^{N(i)} (\mathbf{x}_{n_k} - \mathbf{x}_i) w_{in_k} = \sum_{j \in \mathcal{N}(i)} \mathbf{x}_j w_{ij} - \left[ \sum_{j \in \mathcal{N}(i)} w_{ij} \right] \mathbf{x}_i \\ &= \sum_{j \in \mathcal{N}(i)} \mathbf{x}_j w_{ij} - \mathbf{x}_i, \end{aligned}$$

so that

$$E_{\text{loc}}^{(i)}(\tilde{\mathbf{w}}_i) = \|\tilde{\mathbf{X}}_i \tilde{\mathbf{w}}_i\|^2.$$

Finding the optimal  $\tilde{\mathbf{w}}_i$  that minimises the error  $E_{\text{loc}}^{(i)}$  subject to (1) reduces to finding a critical point of the Lagrange function

$$\mathcal{L}_i = \frac{1}{2} \|\tilde{\mathbf{X}}_i \tilde{\mathbf{w}}_i\|^2 - \lambda_i \left[ \sum_{j \in \mathcal{N}(i)} w_{ij} - 1 \right],$$

where  $\lambda_i$  is a scalar Lagrange multiplier. The latter task, in turn, requires the computation of the derivative  $\delta \mathcal{L}_i / \delta \tilde{\mathbf{w}}_i$  of  $\mathcal{L}_i$  with respect to  $\tilde{\mathbf{w}}_i$ . For each  $m = 1, 2, \dots$ , denote by  $\mathbf{1}_m$  the length- $m$  vector of all ones. It turns out that  $\delta \mathcal{L}_i / \delta \tilde{\mathbf{w}}_i$  can be represented as the length- $N(i)$  vector given by

$$\frac{\delta \mathcal{L}_i}{\delta \tilde{\mathbf{w}}_i} = \tilde{\mathbf{X}}_i^\top \tilde{\mathbf{X}}_i \tilde{\mathbf{w}}_i - \lambda_i \mathbf{1}_{N(i)}$$

(see Appendix A). Here  $\tilde{\mathbf{X}}_i^\top \tilde{\mathbf{X}}_i$  has the meaning of the *local covariance* matrix associated with  $\mathbf{x}_i$  and will henceforth be denoted by  $\tilde{\mathbf{C}}_i$ . The optimal weight  $\tilde{\mathbf{w}}_i$  satisfies the system comprising the equation  $\delta \mathcal{L}_i / \delta \tilde{\mathbf{w}}_i = \mathbf{0}$  and condition (1), that is,

$$\tilde{\mathbf{C}}_i \tilde{\mathbf{w}}_i - \lambda_i \mathbf{1}_{N(i)} = \mathbf{0} \quad \text{and} \quad \mathbf{1}_{N(i)}^\top \tilde{\mathbf{w}}_i = 1.$$

The first of these equations implies that  $\tilde{\mathbf{w}}_i = \lambda_i \tilde{\mathbf{C}}_i^{-1} \mathbf{1}_{N(i)}$ , and this equality coupled with the second equation of the above system yields  $\lambda_i \mathbf{1}_{N(i)}^\top \tilde{\mathbf{C}}_i^{-1} \mathbf{1}_{N(i)} = 1$  and further  $\lambda_i = (\mathbf{1}_{N(i)}^\top \tilde{\mathbf{C}}_i^{-1} \mathbf{1}_{N(i)})^{-1}$ . Hence, finally,

$$\tilde{\mathbf{w}}_i = (\mathbf{1}_{N(i)}^\top \tilde{\mathbf{C}}_i^{-1} \mathbf{1}_{N(i)})^{-1} \tilde{\mathbf{C}}_i^{-1} \mathbf{1}_{N(i)}.$$

#### 4. Searching for the Optimal Embedding

An embedding of the data set  $\mathbf{x}_1, \dots, \mathbf{x}_N$  into the Euclidean space of dimension  $d$  is a mapping which assigns to each  $\mathbf{x}_i$  a vector  $\mathbf{y}_i$  in  $\mathbb{R}^d$ . The embedding vectors can be compactly represented by the  $d \times N$  matrix  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]$ . The embedding matrix sought by LLE minimises the function

$$E_{\text{lle}}(\mathbf{Y}) = \sum_{i=1}^N \left\| \mathbf{y}_i - \sum_{j=1}^N w_{ij} \mathbf{y}_j \right\|^2$$

subject to the conditions:

$$\sum_{i=1}^N \mathbf{y}_i = \mathbf{0}, \quad (3)$$

$$N^{-1} \sum_{i=1}^N \mathbf{y}_i \mathbf{y}_i^\top = \mathbf{I}_d, \quad (4)$$

where  $\mathbf{I}_d$  denotes the  $d \times d$  unit matrix. The imposing of constraints (3) and (4) guarantees the essential uniqueness of the solution (up to an orthogonal transformation of the embedding space; see below). The first condition removes the translational degree of freedom related to the fact that  $E_{\text{lle}}$  is unaltered when each  $\mathbf{y}_i$  is replaced by  $\mathbf{y}_i + \mathbf{t}$ , where  $\mathbf{t}$  is a length- $d$  vector. The second condition ensures that reconstruction errors for different co-ordinates in the embedding are measured on the same scale.

##### 4.1. Equivalent formulation

Noting that  $\sum_{j=1}^N w_{ij} \mathbf{y}_j$  is the  $i$ th column of  $\mathbf{Y} \mathbf{W}^\top$  for each  $1 \leq i \leq N$  and that

$$\|\mathbf{Y} - \mathbf{Y} \mathbf{W}^\top\|_F^2 = \sum_{i=1}^N \left\| \mathbf{y}_i - \sum_{j=1}^N w_{ij} \mathbf{y}_j \right\|^2,$$

with  $\|\cdot\|_F$  the Frobenius norm, leads to

$$E_{\text{lle}}(\mathbf{Y}) = \|\mathbf{Y} - \mathbf{Y} \mathbf{W}^\top\|_F^2 \quad (5)$$

as an equivalent expression for the embedding cost function. Conditions (3) and (4) can equivalently be formulated as

$$\mathbf{Y} \mathbf{1}_N = \mathbf{0}, \quad (6)$$

$$N^{-1} \mathbf{Y} \mathbf{Y}^\top = \mathbf{I}_d. \quad (7)$$

It follows from (5) that the value of  $E_{\text{lle}}$  does not change if  $\mathbf{Y}$  is replaced by  $\mathbf{UY}$ , where  $\mathbf{U}$  is a member of the set  $O(d)$  of all real orthogonal  $d \times d$  matrices, which is a group under matrix multiplication. Likewise constraints (6) and (7) are invariant to premultiplication by a matrix  $\mathbf{U}$  in  $O(d)$ , as  $\mathbf{Y}\mathbf{1}_N = \mathbf{0}$  implies  $\mathbf{UY}\mathbf{1}_N = \mathbf{0}$  and  $N^{-1}\mathbf{Y}\mathbf{Y}^\top = \mathbf{I}_d$  implies  $N^{-1}\mathbf{UY}\mathbf{Y}^\top\mathbf{U}^\top = \mathbf{U}\mathbf{I}_d\mathbf{U}^\top = \mathbf{U}\mathbf{U}^\top = \mathbf{I}_d$ . Thus the solution to the problem of the constrained minimisation of  $E_{\text{lle}}$  is determined only up to a left factor in the form of a  $d \times d$  orthogonal matrix. The search space for this problem is effectively the set of all equivalence classes of  $d \times N$  matrices, with two matrices belonging to the same class if one can be obtained from the other by premultiplication by a  $d \times d$  orthogonal matrix. It turns out that this search space can be usefully reduced, with the resulting smaller search space naturally identified as an appropriate Grassmann manifold.

Indeed, consider the ‘tall and skinny’ matrix  $\mathbf{Y}^\top$  instead of the ‘short and fat’ matrix  $\mathbf{Y}$ . Then condition (7) becomes the requirement that  $N^{-1/2}\mathbf{Y}^\top$  should belong to the  $(N, d)$ -Stiefel manifold  $\text{St}(N, d)$ . By definition, the  $(N, d)$ -Stiefel manifold is the set of all  $N \times d$  orthogonal matrices

$$\text{St}(N, d) = \{\mathbf{X} \in \mathbb{R}^{N \times d} \mid \mathbf{X}^\top\mathbf{X} = \mathbf{I}_d\}$$

(see Refs. 1 and 3). Note that when  $d = 1$ , this set is the  $(N - 1)$ -dimensional unit sphere. The original problem of minimising  $E_{\text{lle}}$  on  $\mathbb{R}^{d \times N}$  subject to (6) and (7) can be cast as minimising  $E_{\text{lle}}$  on  $\text{St}(N, d)$  subject to (6) as a single constraint. The solution of this new problem is still ambiguous: if  $N^{-1/2}\mathbf{Y}^\top$  represents a particular solution, then each matrix of the form  $N^{-1/2}\mathbf{Y}^\top\mathbf{U}$ , where  $\mathbf{U} \in O(d)$ , represents an essentially equivalent solution. The set of all such equivalent solutions

$$[N^{-1/2}\mathbf{Y}^\top] = \{N^{-1/2}\mathbf{Y}^\top\mathbf{U} \mid \mathbf{U} \in O(d)\},$$

termed the  $O(d)$ -orbit of  $N^{-1/2}\mathbf{Y}^\top$ , determines a single member of the  $(N, d)$ -Grassmann manifold  $\text{Gr}(N, d)$ . By definition, this latter manifold is the quotient space of the  $(N, d)$ -Stiefel manifold obtained by declaring  $\mathbf{X}, \mathbf{Y} \in \text{St}(N, d)$  to be equivalent if there exists a  $d \times d$  orthogonal matrix  $\mathbf{U}$  such that  $\mathbf{Y} = \mathbf{X}\mathbf{U}$ . Now,  $E_{\text{lle}}$  induces a cost function on  $\text{Gr}(N, d)$

$$\tilde{E}_{\text{lle}}([\mathbf{X}]) = E_{\text{lle}}(\mathbf{X}) \quad (\mathbf{X} \in \text{St}(N, d))$$

and likewise (6) induces a constraint on  $\text{Gr}(N, d)$  simultaneously satisfied or not satisfied by all elements of any given  $O(d)$ -orbit. With these induced objects in place, the corresponding constrained minimiser for  $\tilde{E}_{\text{lle}}$  can be viewed as an entity from which the ambiguity related to the  $O(d)$ -invariance property of the initial optimisation problem is ‘‘quotiented out’’.

As a set of points, the  $(N, d)$ -Grassmann manifold can be identified with the set of all  $d$ -dimensional linear subspaces of  $\mathbb{R}^N$ . Under this identification, any particular  $\mathbf{X} \in \text{St}(N, d)$  represents an orthonormal basis for the  $d$ -dimensional subspace in  $\text{Gr}(N, d)$  spanned by the columns of  $\mathbf{X}$ , and any  $d \times d$  orthogonal  $\mathbf{U}$  corresponds to a change-of-basis transformation, namely, the one that takes the

orthonormal basis represented by  $\mathbf{X}$  to the new orthonormal basis, for the same subspace, represented by  $\mathbf{X}\mathbf{U}$ . When  $d = 1$ , the  $(N, d)$ -Grassmann manifold reduces to the  $(N - 1)$ -dimensional projective space. The  $(N, d)$ -Grassmann manifold can be equipped with a natural topology under which it becomes a compact space. The compactness property of  $\text{Gr}(N, d)$  guarantees the existence of a solution of the constrained optimisation problem for  $\tilde{E}_{\text{lle}}$ , and hence a (non-univocal) solution of the original constrained optimisation problem for  $E_{\text{lle}}$ .

#### 4.2. Computing the optimal embedding

To isolate the optimal embedding matrix  $\mathbf{Y}$  (to within  $O(d)$ -equivalence), introduce the Lagrange function

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \sum_{i=1}^N \left\| \mathbf{y}_i - \sum_{j=1}^N \mathbf{y}_j w_{ij} \right\|^2 - \frac{1}{2} \sum_{\alpha, \beta=1}^d \lambda_{\alpha\beta} \left[ \frac{1}{N} \sum_{i=1}^N y_{\alpha i} y_{\beta i} - \delta_{\alpha\beta} \right] \\ & - \sum_{\alpha=1}^d \kappa_{\alpha} \left[ \frac{1}{N} \sum_{i=1}^N y_{\alpha i} \right], \end{aligned}$$

where  $\mathbf{y}_i = [y_{1i}, \dots, y_{di}]^T$ , the  $\lambda_{\alpha\beta}$  satisfying  $\lambda_{\alpha\beta} = \lambda_{\beta\alpha}$  form one group of scalar Lagrange multipliers, the  $\kappa_{\alpha}$  form another group of scalar Lagrange multipliers, and the  $\delta_{\alpha\beta}$  are components of the Kronecker delta, with  $\delta_{\alpha\beta} = 0$  if  $\alpha \neq \beta$  and  $\delta_{\alpha\beta} = 1$  if  $\alpha = \beta$ . On letting  $\mathbf{\Lambda} = [\lambda_{\alpha\beta}]_{1 \leq \alpha, \beta \leq d} \in \mathbb{R}^{d \times d}$  and  $\boldsymbol{\kappa} = [\kappa_1, \dots, \kappa_d]^T \in \mathbb{R}^{d \times 1}$ , this function can be written as

$$\mathcal{L} = \frac{1}{2} \|\mathbf{Y} - \mathbf{Y}\mathbf{W}^T\|_F^2 - \frac{1}{2} \text{tr} \left( \mathbf{\Lambda} \left[ \frac{1}{N} \mathbf{Y}\mathbf{Y}^T - \mathbf{I}_d \right] \right) - \frac{1}{N} \mathbf{1}_N^T \mathbf{Y} \boldsymbol{\kappa},$$

where  $\text{tr}$  denotes trace. With the derivative  $\delta\mathcal{L}/\delta\mathbf{Y}$  of  $\mathcal{L}$  with respect to  $\mathbf{Y}$  represented by the  $d \times N$  matrix as per

$$\frac{\delta\mathcal{L}}{\delta\mathbf{Y}} = \mathbf{Y}(\mathbf{I}_N - \mathbf{W}^T)(\mathbf{I}_N - \mathbf{W}) - N^{-1} \mathbf{\Lambda} \mathbf{Y} - \boldsymbol{\kappa} \mathbf{1}_N^T$$

(see Appendix B), the optimal  $\mathbf{Y}$  satisfies

$$\mathbf{Y}(\mathbf{I}_N - \mathbf{W}^T)(\mathbf{I}_N - \mathbf{W}) - N^{-1} \mathbf{\Lambda} \mathbf{Y} - \boldsymbol{\kappa} \mathbf{1}_N^T = \mathbf{0} \quad (8)$$

together with the feasibility conditions (6) and (7). In the overall system comprising (6), (7), and (8), the second equation reduces the search space to the  $(N, d)$ -Stiefel manifold, and the remaining equations furnish constraints on this manifold. The solution of the system is to be determined to within  $O(d)$ -equivalence.

Equation (8) can be simplified as follows. First, by postmultiplying (8) by  $\mathbf{1}_N$ ,

$$\mathbf{Y}(\mathbf{I}_N - \mathbf{W}^T)(\mathbf{I}_N - \mathbf{W})\mathbf{1}_N - N^{-1} \mathbf{\Lambda} \mathbf{Y} \mathbf{1}_N - \boldsymbol{\kappa} \mathbf{1}_N^T \mathbf{1}_N = \mathbf{0}. \quad (9)$$

Next, by conditions (1) and (2) relating to all indices  $1 \leq i \leq N$ ,

$$(\mathbf{I}_N - \mathbf{W})\mathbf{1}_N = \mathbf{0}. \quad (10)$$

The last equality implies that the first term of (9) is null. In view of (6), the second term of (9) is also null. As a result, the third term of (9) has to vanish, and, given that  $\mathbf{1}_N^\top \mathbf{1}_N = N$ , it follows that  $\boldsymbol{\kappa} = 0$ . Thus (8) reduces to

$$\mathbf{Y}(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W}) = N^{-1}\boldsymbol{\Lambda}\mathbf{Y},$$

or, equivalently, by taking the transpose of both sides of the equation, to

$$(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W})\mathbf{Y}^\top = N^{-1}\mathbf{Y}^\top\boldsymbol{\Lambda}. \quad (11)$$

At this point, it is critical to note that, with  $\boldsymbol{\Lambda}$  and  $\mathbf{Y}$  treated as dependent unknowns, the above equation has a special symmetry — it is invariant to an action of  $O(d)$ . Specifically, if a pair  $(\boldsymbol{\Lambda}, \mathbf{Y})$  satisfies (11), then so too does the pair  $(\tilde{\boldsymbol{\Lambda}}, \tilde{\mathbf{Y}}) = (\mathbf{U}^\top\boldsymbol{\Lambda}\mathbf{U}, \mathbf{U}^\top\mathbf{Y})$  for any  $\mathbf{U} \in O(d)$ . Any particular selection of  $\mathbf{U}$  defines a *gauge*, the passage from  $(\boldsymbol{\Lambda}, \mathbf{Y})$  to  $(\tilde{\boldsymbol{\Lambda}}, \tilde{\mathbf{Y}})$  for some  $\mathbf{U}$  is an instance of a *gauge transformation*, and the existence of equivalent representations of  $(\boldsymbol{\Lambda}, \mathbf{Y})$  as  $(\tilde{\boldsymbol{\Lambda}}, \tilde{\mathbf{Y}})$  corresponds to the *gauge freedom*. The gauge freedom in (11) is a direct reflection of the fact that a “true” search space for the optimisation problem under study is not the Stiefel manifold  $\text{St}(N, d)$  but rather the Grassmann manifold  $\text{Gr}(N, d)$ .

The symmetry of (11) can now be exploited by a *gauge fixing* that consists in a choice of a particularly convenient gauge. Suppose  $(\boldsymbol{\Lambda}, \mathbf{Y})$  is a specific solution to the system consisting of (6), (7), and (11). Let  $\boldsymbol{\Lambda} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$  be the eigenvalue decomposition of  $\boldsymbol{\Lambda}$ , with  $\mathbf{U}$  a  $d \times d$  orthogonal matrix and  $\mathbf{D} = \text{diag}(d_1, \dots, d_d)$ . Rearrange the diagonal entries of  $\mathbf{D}$  in increasing order by using an appropriate  $d \times d$  permutation matrix<sup>a</sup>  $\mathbf{P}$  as follows

$$\mathbf{P}^\top\mathbf{D}\mathbf{P} = \mathbf{D}' = \text{diag}(d'_1, \dots, d'_d) \quad \text{with} \quad d'_1 \leq \dots \leq d'_d.$$

Note that  $\mathbf{P}$ , as any other permutation matrix, is orthogonal, and so too is the composite matrix  $\mathbf{U}\mathbf{P}$ . Adopting the gauge associated with  $\mathbf{U}\mathbf{P}$  leads to the gauge transformation  $(\tilde{\boldsymbol{\Lambda}}, \tilde{\mathbf{Y}}) = (\mathbf{D}', \mathbf{P}^\top\mathbf{U}^\top\mathbf{Y})$ . Under this transformation (11) becomes

$$(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W})\tilde{\mathbf{Y}}^\top = N^{-1}\tilde{\mathbf{Y}}^\top\mathbf{D}'.$$

The last equation shows that the columns of  $\tilde{\mathbf{Y}}^\top$  (or, equivalently, the rows of  $\tilde{\mathbf{Y}}$ ) are eigenvectors of  $(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W})$ . More specifically, if  $\mathbf{v}_i$  is the  $i$ th column vector of  $\tilde{\mathbf{Y}}^\top$ , then

$$(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W})\mathbf{v}_i = \lambda_i\mathbf{v}_i, \quad \lambda_i = N^{-1}d'_i.$$

Given that  $N^{-1}\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\top = \mathbf{I}_d$ , each  $\mathbf{v}_i$  has norm  $N^{1/2}$ . Thus the columns of  $N^{-1/2}\tilde{\mathbf{Y}}$  (that is, the  $N^{-1/2}\mathbf{v}_i$ ) are normalised eigenvectors of  $(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W})$ . It now remains to find out precisely which of all the normalised eigenvectors of  $(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W})$  are those forming the columns of  $N^{-1/2}\tilde{\mathbf{Y}}$ .

<sup>a</sup>A permutation matrix is a square matrix whose entries are all 0's and 1's, with exactly one 1 in each row and exactly one 1 in each column. Premultiplying an  $n \times n$  matrix  $A$  by an  $n \times n$  permutation matrix results in a rearrangement of the rows of  $A$ . Postmultiplying an  $n \times n$  matrix  $A$  by an  $n \times n$  permutation matrix results in a rearrangement of the columns of  $A$ .

It follows from (5) that

$$E_{\text{lle}}(\tilde{\mathbf{Y}}) = \text{tr}(\tilde{\mathbf{Y}}(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W})\tilde{\mathbf{Y}}^\top)$$

and from (7) and (11) that

$$\tilde{\mathbf{Y}}(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W})\tilde{\mathbf{Y}}^\top = \mathbf{D}'.$$

Hence

$$E_{\text{lle}}(\tilde{\mathbf{Y}}) = \text{tr} \mathbf{D}' = N \sum_{i=1}^d \lambda_i. \quad (12)$$

Because  $(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W})$  is positive semi-definite, each  $\lambda_i$  is non-negative. Now, by (10),  $\mathbf{1}_N$  is an eigenvector of  $(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W})$  with zero eigenvalue. Since (6) can be rewritten as  $\mathbf{1}_N^\top \mathbf{Y}^\top = \mathbf{0}^\top$ , it follows that

$$\mathbf{1}_N^\top \tilde{\mathbf{Y}}^\top = \mathbf{1}_N^\top \mathbf{Y}^\top \mathbf{U} \mathbf{P} = \mathbf{0}^\top,$$

showing that  $\mathbf{1}_N$  is orthogonal to each  $\mathbf{v}_i$ . Generically,  $\mathbf{1}_N$  is the only (up to scale) eigenvector of  $(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W})$  with zero eigenvalue, so each  $\lambda_i$  can be assumed positive. Since, again generically, the  $\lambda_i$  are also all distinct, it further follows that each  $N^{-1/2}\mathbf{v}_i$  can be assumed to be uniquely specified (to within the sign ambiguity) by  $\lambda_i$ . Now, in view of (12), the  $N^{-1/2}\mathbf{v}_i$  are characterised as those normalised eigenvectors of  $(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W})$  for which  $\sum_{i=1}^d \lambda_i$  attains the smallest possible value. But, as the  $\lambda_i$  all differ, the smallest value of  $\sum_{i=1}^d \lambda_i$  is attained precisely when  $\lambda_i$  is the  $i$ th smallest *positive* eigenvalue of  $(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W})$  for each  $1 \leq i \leq d$ . Because the smallest eigenvalue of  $(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W})$  is zero, the  $i$ th smallest *positive* eigenvalue of  $(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W})$  coincides with the  $(i+1)$ th smallest eigenvalue of  $(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W})$ . Consequently, the optimal solution expressed in terms of  $\tilde{\mathbf{Y}}$  is characterised by the requirement that, for each  $1 \leq i \leq d$ , the  $i$ th column vector of  $N^{-1/2}\tilde{\mathbf{Y}}^\top$  be a normalised eigenvector of  $(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W})$  corresponding to the  $(i+1)$ th smallest eigenvalue, with any of the two candidate eigenvectors considered as equally valid if they differ merely by a sign. The embedding matrix  $\tilde{\mathbf{Y}}$  satisfying this condition is precisely what LLE takes as an output. It is well worth stressing that  $\tilde{\mathbf{Y}}$  is a particular solution differing from the starting  $\mathbf{Y}$  by a left factor embodied by  $d \times d$  orthogonal matrix. All equivalent solutions take the form  $\mathbf{U}\tilde{\mathbf{Y}}$  with  $\mathbf{U}$  varying over  $O(d)$ .

## 5. Concluding Remarks

The optimisation problem solved by the LLE algorithm has been presented as one posed on a Grassmann manifold. The advantage of this formulation is at least twofold. First, it helps reveal the internal degrees of freedom present in the natural, coordinate-specific formulation of the problem that initially involves a Euclidean space and upon refinement operates with a Stiefel manifold. Second, with the gauge freedom properly recognised, it facilitates an adequate choice of a coordinate system for reducing the equation for a critical point of the Lagrange function, in terms of which the problem is expressed, to an easily solvable eigenvalue problem.

### Acknowledgements

This work was supported by the Australian Research Council.

### Appendix A. Calculation of the Derivative of $\mathcal{L}_i$

For a scalar function  $f(\tilde{\mathbf{w}})$  with a vector argument  $\tilde{\mathbf{w}}$ , the derivative of  $f$  can be defined as the unique (column) vector  $\delta f/\delta\tilde{\mathbf{w}}$  of the same length as  $\tilde{\mathbf{w}}$ , satisfying

$$\begin{aligned} f(\tilde{\mathbf{w}} + \delta\tilde{\mathbf{w}}) &= f(\tilde{\mathbf{w}}) + \left(\frac{\delta f}{\delta\tilde{\mathbf{w}}}\right)^\top \delta\tilde{\mathbf{w}} + o(\|\delta\tilde{\mathbf{w}}\|) \\ &= f(\tilde{\mathbf{w}}) + \text{tr}\left(\frac{\delta f}{\delta\tilde{\mathbf{w}}} \delta\tilde{\mathbf{w}}^\top\right) + o(\|\delta\tilde{\mathbf{w}}\|), \end{aligned}$$

where  $o(\epsilon)$  stands for a term that approaches zero faster than  $\epsilon$  as  $\epsilon \rightarrow 0$ .

To obtain an explicit expression for  $\delta\mathcal{L}_i/\delta\tilde{\mathbf{w}}_i$ , we first calculate the derivative of the function

$$g(\tilde{\mathbf{w}}) = \frac{1}{2}\|\tilde{\mathbf{X}}\tilde{\mathbf{w}}\|^2.$$

We shall use elementary matrix-calculation rules such as: (i)  $\text{tr}(\alpha) = \alpha$  if  $\alpha$  is scalar; (ii)  $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^\top)$ ; and (iii)  $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$ . Note that

$$\|\tilde{\mathbf{X}}\tilde{\mathbf{w}}\|^2 = \text{tr}(\tilde{\mathbf{X}}\tilde{\mathbf{w}}\tilde{\mathbf{w}}^\top\tilde{\mathbf{X}}^\top)$$

and

$$\begin{aligned} \|\tilde{\mathbf{X}}(\tilde{\mathbf{w}} + \delta\tilde{\mathbf{w}})\|^2 &= \text{tr}(\tilde{\mathbf{X}}\tilde{\mathbf{w}}\tilde{\mathbf{w}}^\top\tilde{\mathbf{X}}^\top) + \text{tr}(\tilde{\mathbf{X}}\delta\tilde{\mathbf{w}}\tilde{\mathbf{w}}^\top\tilde{\mathbf{X}}^\top) \\ &\quad + \text{tr}(\tilde{\mathbf{X}}\tilde{\mathbf{w}}\delta\tilde{\mathbf{w}}^\top\tilde{\mathbf{X}}^\top) + \text{tr}(\tilde{\mathbf{X}}\delta\tilde{\mathbf{w}}\delta\tilde{\mathbf{w}}^\top\tilde{\mathbf{X}}^\top). \end{aligned}$$

Now

$$\text{tr}(\tilde{\mathbf{X}}\delta\tilde{\mathbf{w}}\tilde{\mathbf{w}}^\top\tilde{\mathbf{X}}^\top) = \text{tr}(\delta\tilde{\mathbf{w}}\tilde{\mathbf{w}}^\top\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}) = \text{tr}(\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}\tilde{\mathbf{w}}\delta\tilde{\mathbf{w}}^\top)$$

and

$$\text{tr}(\tilde{\mathbf{X}}\tilde{\mathbf{w}}\delta\tilde{\mathbf{w}}^\top\tilde{\mathbf{X}}^\top) = \text{tr}(\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}\tilde{\mathbf{w}}\delta\tilde{\mathbf{w}}^\top).$$

Hence

$$g(\tilde{\mathbf{w}} + \delta\tilde{\mathbf{w}}) - g(\tilde{\mathbf{w}}) = \text{tr}(\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}\tilde{\mathbf{w}}\delta\tilde{\mathbf{w}}^\top) + O(\|\delta\tilde{\mathbf{w}}\|^2),$$

where  $O(\epsilon)$  denotes a term that is bounded by a constant factor of  $\epsilon$  near 0, leading to

$$\frac{\delta g}{\delta\tilde{\mathbf{w}}} = \tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}\tilde{\mathbf{w}}. \tag{A.1}$$

We now calculate the derivative of the function

$$\varphi(\tilde{\mathbf{w}}) = \sum_{n=1}^m w_n - 1 = \mathbf{1}_m^\top \tilde{\mathbf{w}} - 1$$

with  $\tilde{\mathbf{w}} = [w_1, \dots, w_m]^\top$ . In view of  $\mathbf{1}_m^\top \tilde{\mathbf{w}} = \text{tr}(\mathbf{1}_m \tilde{\mathbf{w}}^\top)$ , we have

$$\varphi(\tilde{\mathbf{w}} + \delta\tilde{\mathbf{w}}) - \varphi(\tilde{\mathbf{w}}) = \text{tr}(\mathbf{1}_m \delta\tilde{\mathbf{w}}^\top)$$

whence, immediately,

$$\frac{\delta\varphi}{\delta\tilde{\mathbf{w}}} = \mathbf{1}_m. \quad (\text{A.2})$$

Applying (A.1) and (A.2) to the respective terms of  $\mathcal{L}_i$  finally gives

$$\frac{\delta\mathcal{L}_i}{\delta\tilde{\mathbf{w}}_i} = \tilde{\mathbf{X}}_i^\top \tilde{\mathbf{X}}_i \tilde{\mathbf{w}}_i - \lambda_i \mathbf{1}_{N(i)}.$$

### Appendix B. Calculation of the Derivative of $\mathcal{L}$

For a scalar function  $f = f(\mathbf{Y})$  with a matrix argument  $\mathbf{Y}$ , the derivative of  $f$  can be defined as the unique matrix  $\delta f / \delta \mathbf{Y}$  of the same size as  $\mathbf{Y}$ , satisfying

$$f(\mathbf{Y} + \delta\mathbf{Y}) = f(\mathbf{Y}) + \text{tr}\left(\frac{\delta f}{\delta \mathbf{Y}} \delta\mathbf{Y}^\top\right) + o(\|\delta\mathbf{Y}\|_F).$$

To calculate  $\delta\mathcal{L}/\delta\mathbf{Y}$ , first set  $\mathbf{S} = \mathbf{I}_N - \mathbf{W}^\top$  and let

$$f_1(\mathbf{W}) = \frac{1}{2} \|\mathbf{Y} - \mathbf{Y}\mathbf{W}^\top\|_F^2 = \frac{1}{2} \|\mathbf{Y}\mathbf{S}\|_F^2.$$

Combining  $\|\mathbf{Y}\mathbf{S}\|_F^2 = \text{tr}(\mathbf{Y}\mathbf{S}\mathbf{S}^\top\mathbf{Y}^\top)$  and

$$\begin{aligned} \|(\mathbf{Y} + \delta\mathbf{Y})\mathbf{S}\|_F^2 &= \text{tr}(\mathbf{Y}\mathbf{S}\mathbf{S}^\top\mathbf{Y}^\top) + \text{tr}(\delta\mathbf{Y}\mathbf{S}\mathbf{S}^\top\mathbf{Y}^\top) \\ &\quad + \text{tr}(\mathbf{Y}\mathbf{S}\mathbf{S}^\top\delta\mathbf{Y}^\top) + \text{tr}(\delta\mathbf{Y}\mathbf{S}\mathbf{S}^\top\delta\mathbf{Y}^\top) \end{aligned}$$

with  $\text{tr}(\delta\mathbf{Y}\mathbf{S}\mathbf{S}^\top\mathbf{Y}^\top) = \text{tr}(\mathbf{Y}\mathbf{S}\mathbf{S}^\top\delta\mathbf{Y}^\top)$  yields

$$f_1(\mathbf{Y} + \delta\mathbf{Y}) - f_1(\mathbf{Y}) = \text{tr}(\mathbf{Y}\mathbf{S}\mathbf{S}^\top\delta\mathbf{Y}^\top) + O(\|\delta\mathbf{Y}\|_F^2),$$

whence

$$\frac{\delta f_1}{\delta \mathbf{Y}} = \mathbf{Y}\mathbf{S}\mathbf{S}^\top = \mathbf{Y}(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W}).$$

Define

$$f_2(\mathbf{Y}) = \frac{1}{2} \text{tr}\left(\mathbf{\Lambda}\left[\frac{1}{N}\mathbf{Y}\mathbf{Y}^\top - \mathbf{I}_d\right]\right).$$

Since

$$\begin{aligned} \text{tr}(\mathbf{\Lambda}(\mathbf{Y} + \delta\mathbf{Y})(\mathbf{Y} + \delta\mathbf{Y})^\top) &= \text{tr}(\mathbf{\Lambda}\mathbf{Y}\mathbf{Y}^\top) + \text{tr}(\mathbf{\Lambda}\delta\mathbf{Y}\mathbf{Y}^\top) \\ &\quad + \text{tr}(\mathbf{\Lambda}\mathbf{Y}\delta\mathbf{Y}^\top) + \text{tr}(\mathbf{\Lambda}\delta\mathbf{Y}\delta\mathbf{Y}^\top) \end{aligned}$$

and

$$\text{tr}(\mathbf{\Lambda}\delta\mathbf{Y}\mathbf{Y}^\top) = \text{tr}(\delta\mathbf{Y}\mathbf{Y}^\top\mathbf{\Lambda}) = \text{tr}(\mathbf{\Lambda}^\top\mathbf{Y}\delta\mathbf{Y}^\top) = \text{tr}(\mathbf{\Lambda}\mathbf{Y}\delta\mathbf{Y}^\top),$$

where the last equality uses the fact that  $\mathbf{\Lambda}$  is symmetric, it follows that

$$f_2(\mathbf{Y} + \delta\mathbf{Y}) - f_2(\mathbf{Y}) = \frac{1}{N} \text{tr}(\mathbf{\Lambda}\mathbf{Y}\delta\mathbf{Y}^\top) + O(\|\delta\mathbf{Y}\|_F^2),$$

and hence that

$$\frac{\delta f_2}{\delta \mathbf{Y}} = \frac{1}{N} \mathbf{\Lambda} \mathbf{Y}.$$

Now let

$$f_3(\mathbf{Y}) = \mathbf{1}_N^\top \mathbf{Y} \boldsymbol{\kappa}.$$

From the identities

$$\mathbf{1}_N^\top \mathbf{Y} \boldsymbol{\kappa} = \text{tr}(\mathbf{1}_N^\top \mathbf{Y} \boldsymbol{\kappa}) = \text{tr}(\boldsymbol{\kappa} \mathbf{1}_N^\top \mathbf{Y})$$

it follows that

$$f_3(\mathbf{Y} + \delta \mathbf{Y}) - f_3(\mathbf{Y}) = \text{tr}(\boldsymbol{\kappa} \mathbf{1}_N^\top \delta \mathbf{Y}),$$

whence

$$\frac{\delta f_3}{\delta \mathbf{Y}} = \boldsymbol{\kappa} \mathbf{1}_N^\top.$$

Thus, finally,

$$\begin{aligned} \frac{\delta \mathcal{L}}{\delta \mathbf{Y}} &= \frac{\delta f_1}{\delta \mathbf{Y}} - \frac{\delta f_2}{\delta \mathbf{Y}} - \frac{\delta f_3}{\delta \mathbf{Y}} \\ &= \mathbf{Y}(\mathbf{I}_N - \mathbf{W}^\top)(\mathbf{I}_N - \mathbf{W}) - N^{-1} \mathbf{\Lambda} \mathbf{Y} - \boldsymbol{\kappa} \mathbf{1}_N^\top. \end{aligned}$$

## References

1. P.-A. Absil, R. Mahony and R. Sepulchre, Riemannian geometry of Grassmann manifolds with a view on algorithmic computation, *Acta Appl. Math.* **80** (2004) 199–220.
2. C. J. C. Burges, Geometric methods for feature extraction and dimensional reduction: a guided tour, in O. Maimon and L. Rokach (eds.), *The Data Mining and Knowledge Discovery Handbook: A Complete Guide for Research Scientists and Practitioners* (Springer, New York, 2005) pp. 59–92.
3. A. Edelman, T. A. Arias and S. T. Smith, The geometry of algorithms with orthogonality constraints, *SIAM J. Matrix Anal. Appl.* **20** (1998) 303–353.
4. A. Elgammal and C.-S. Lee, Inferring 3D body pose from silhouettes using activity manifold learning, in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Vol. 2 (2004) pp. 681–688.
5. S. Fiori, Quasi-geodesic neural learning algorithms over the orthogonal group: a tutorial, *J. Machine Learning Research* **6** (2005) 743–781.
6. Y. Fu, R. Shen and L. Shen, Digital watermarking based on locally linear embedding, in *Proc. Int. Conf. Image Processing*, Vol. 4 (2004) pp. 2143–2146.
7. T. Han and D. G. Goodenough, Nonlinear feature extraction of hyperspectral data based on locally linear embedding (LLE), in *Proc. IEEE Int. Geoscience and Remote Sensing Symposium*, Vol. 2 (2005) pp. 1237–1240.
8. D. H. Kim and L. H. Finkel, Hyperspectral image processing using locally linear embedding, in *Proc. First Int. IEEE EMBS Conf. Neural Engineering* (2003) pp. 316–319.
9. H.-G. Li, C.-P. Shi and X.-G. Li, LLE based gait recognition, in *Proc. Int. Conf. Machine Learning and Cybernetics*, Vol. 7 (2005) pp. 4516–4521.

10. K. Liu, A. Weissenfeld and J. Ostermann, Parameterization of mouth images by LLE and PCA for image-based facial animation, in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Vol. 5 (2006) pp. 461–464.
  11. S. Pang, D. Kim and S. Y. Bang, Face membership authentication using SVM classification tree generated by membership-based LLE data partition, *IEEE Trans. Neural Networks* **16** (2005) 436–446.
  12. S. T. Roweis and L. K. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science* **290** (2000) 2323–2326.
  13. L. K. Saul and S. T. Roweis, Think globally, fit locally: unsupervised learning of low dimensional manifolds, *J. Machine Learning Research* **4** (2003) 119–155.
  14. L. K. Saul, K. Q. Weinberger, F. Sha, J. Ham and D. D. Lee, Spectral methods for dimensionality reduction, in O. Chapelle, B. Schölkopf and A. Zien (eds.), *Semi-Supervised Learning* (MIT Press, 2006) pp. 279–294.
  15. R. Shi, I. F. Shen and W. Chen, Image denoising through locally linear embedding, in *Proc. Int. Conf. Computer Graphics, Imaging and Vision: New Trends* (2005) pp. 147–152.
  16. J. B. Tenenbaum, V. de Silva and J. C. Langford, A global geometric framework for nonlinear dimensionality reduction, *Science* **290** (2000) 2319–2323.
  17. M. Vlachos, C. Domeniconi, D. Gunopulos, G. Kollios and N. Koudas, Non-linear dimensionality reduction techniques for classification and visualization, in *Proc. Eighth ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining* (2002) pp. 645–651.
  18. J. Wang, C. Zhang and Z. Kou, An analytical mapping for LLE and its applications in multi-pose face synthesis, in *Proc. 14th British Machine Vision Conference* (2003)
  19. J. Yin, D. Hu and Z. Zhou, Manifold learning using growing locally linear embedding, in *Proc. IEEE Symposium Computational Intelligence and Data Mining* (2007) pp. 73–80.
  20. X. Zhang, H.-M. Ma, Y.-S. Liu and C.-X. Gao, An extension of locally linear embedding for pose estimation of 3D object, in *Proc. Int. Conf. Machine Learning and Cybernetics*, Vol. 3 (2007) pp. 1672–1677.
-



**Wojciech Chojnacki** received his M.Sc., Ph.D., and D.Sc. degrees, all in mathematics, in 1976 from Warsaw University, in 1980 from the Polish Academy of Sciences, and in 1988 from Warsaw University, respectively. He is a professor of mathematics in the Faculty of Mathematics and Natural Sciences, the College of Sciences, at Cardinal Stefan Wyszyński University in Warsaw. He is concurrently a senior research fellow in the School of Computer Science at the University of Adelaide working on a range of problems in computer vision. His research interests include differential equations, mathematical foundations of computer vision, functional analysis, and harmonic analysis.



**Michael J. Brooks** received his B.Sc. Hons degree in mathematics in 1975 from Hull University and his M.Sc. and Ph.D. degrees in computer science in 1976 and 1983, respectively, from Essex University. He holds the Chair in Artificial Intelligence within the University of Adelaide's School of Computer Science and is presently Deputy Vice-Chancellor (Research) at the University of Adelaide. His research interests include structure from motion, statistical vision parameter estimation, and video surveillance.