Nonlinear Coordinate Unfolding via Principal Curve Projections with Application to Nonlinear BSS

Deniz Erdogmus[1], Umut Ozertem[1]

[1] Department of CSEE, Oregon Health and Science University
Portland, Oregon, USA
{deniz, ozertemu}@csee.ogi.edu

Abstract. Nonlinear independent components analysis (NICA) is known to be an ill-posed problem when only the independence of the sources are sought. Additional constraints on the distribution of the sources or the structure of the mixing nonlinearity are imposed to achieve a solution that is unique in a suitable sense. In this paper, we present a technique that tackles nonlinear blind source separation (NBSS) as a nonlinear invertible coordinate unfolding problem utilizing a recently developed definition of maximum-likelihood principal curves. The proposition would be applicable most conveniently to independent unimodal source distributions with mixtures that have diminishing second order derivatives along the source axes. Application to multimodal sources would be possible with some modifications that are not discussed in this paper. The ill-posed nature of NBSS is also discussed from a differential geometric perspective in this context.

Keywords: Nonlinear independent component analysis, nonlinear blind source separation, principal curves and surfaces, manifold unfolding, nonlinear coordinate transformation.

1 Introduction

Nonlinear blind source separation (NBSS) is an ill-posed problem that requires various sources of a priori knowledge regarding the joint source distribution and topology of the mixing function. In a landmark paper [1] it has been shown that the square nonlinear independent component analysis (NICA) problem is ill-posed – that, without additional constraints it does not have a unique solution accepting the usual generalized scale and permutation ambiguities of ICA – and various conditions that would force the problem to have a unique solution have been proposed. Traditionally several methods, primarily based on parametric model fitting – especially neural networks of various forms – have been proposed in the literature, while minimizing the usual mutual information inspired independence or separation measures [2,3,4]. The general form of the NBSS problem is

$$\mathbf{x} = \mathbf{f}(\mathbf{s}) \tag{1}$$

where $\mathbf{f}$: $\Re^m \rightarrow \Re^n$ maps the source vector $\mathbf{s} \in \Re^m$ to the measured mixture vector $\mathbf{x} \in \Re^n$. In general, $\mathbf{f}$ is assumed to be a smooth function and $m=n$, since even in this case per

existence theorems, there are infinitely many solutions to this ill-posed inverse estimation problem. In practice, unless further physical motivation is provided, there is no reason to be concerned about finding those solutions which correspond to noninvertible maps **f**. Therefore, limiting the search to the space of invertible **f** makes both theoretical and practical sense, especially since without additional domain information about the mechanisms that generate **x**, a noninvertible **f** would cause information loss that cannot be recovered. From this perspective, the additional uniqueness constraint of bounded source distributions could be viewed as a means of limiting the search to invertible **f**, whose domain becomes the finite support of the source density; while the extension of **f** to the whole real-valued vector space might not be invertible, the restriction to the support of $p(\mathbf{s})$, the source distribution could be.

An extensively studied invertible nonlinear mapping topology is the so-called post-nonlinear mixture model of the form $\mathbf{x} = \mathbf{g}(\mathbf{As})$, where **A** introduces an invertible (or full column rank if $n>m$) linear mixture and **g** is an invertible nonlinear map that acts elementwise (only monotonic scaling, not mixing further) on the mixed intermediate vector [4]. This problem is well defined and relatively straightforward extensions of traditional linear ICA techniques are able to cope with the specific mixture topology. An interesting contribution that focused on the overdetermined post-nonlinear mixture problem by exploiting the geodesic distances on nonlinear manifolds was made by Lee et al. [5]. Noticing that the monotonic nonlinearity simply acts as a local metric distortion independently on each coordinate axis, ideas from isometric dimensionality reduction and manifold learning were employed to *unwrap* the nonlinear mixture data manifold into the original source distribution space. Harmeling et al also followed a similar manifold unfolding approach using kernel techniques [6]. The current paper could be considered as an extension of these works focusing on square mixture scenarios where the intrinsic dimensionality of the mixture manifold is identical to that of the sources. The method would be immediately applicable to overdetermined situations where the data lies on a low dimensional manifold embedded in a higher dimensional Euclidean space – the principal coordinates proposed in the following would simply reduce to zeros for any additional dimensions outside the data manifold automatically.

The primary tool that will be exploited in this paper is a recently proposed definition of principal curves and surfaces that follow the maximum likelihood estimation principle, as opposed to the commonly employed conditional least-squares type principal curve/surface/manifold techniques. This new definition provides a well-defined principal manifold structure that underlies any multidimensional probability density function, leading to gradient ascent, mean-shift, or expectation maximization type algorithms for manifold learning.

2 Subspace Maximum Likelihood Principal Manifolds

Principal manifolds are underlying geometrical structures of probability distributions that manifest canonical solutions for denoising and dimensionality reduction. Traditionally, self-consistent principal surfaces defined by Hastie [7] and studied by various researchers [8-12] have relied on the conditional expectation and least-squares

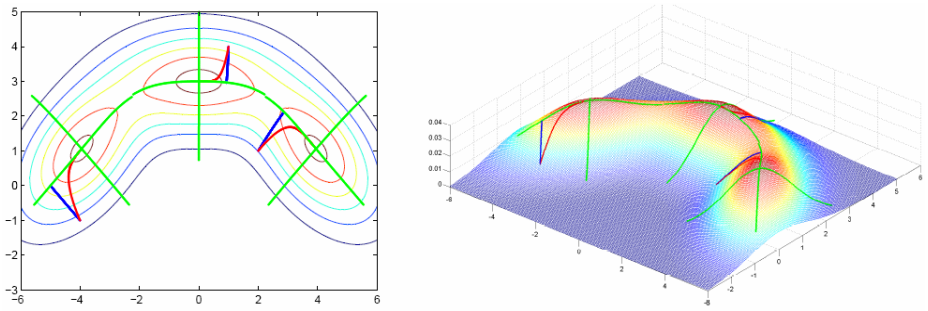

Fig. 1. The 1-dimensional principal manifolds (green) of a 2-dimensional 3-component Gaussian mixture model are shown. The regular gradient ascent trajectories (red) and subspace gradient ascent trajectories (blue) to identify local maxima and subspace local maxima starting from three illustrative points in the space are also shown to emphasize the difference between converging to the usual local maxima and projecting a point to the principal curve.

reconstruction error minimization approach due to the traditional appeal of second-order statistical optimality criteria and the uniqueness of the solution under the self-consistency conditions. This definition creates various practical difficulties for algorithmic solutions to identify such manifolds, besides the theoretical shortcoming that self-intersecting or nonsmooth manifolds are not

Table 1. Generalized local first and second order conditions for *d*-dimensional principal manifolds embedded in *n*-dimensions.

x is a local max iff	**x** is in *d*-dim principal manifold iff
Gradient is zero	Gradient $\perp$ (*n*-*d*) eigenvectors of Hessian
Hessian eigenvalues < 0	Hessian eigenvalues of $\perp$ eigenvectors < 0

acceptable in this framework. Recently, we have proposed a local subspace maxima approach to defining and identifying principal surfaces [13]. This new definition generalizes the usual first and second order derivative conditions to identify local maxima and provides a geometrically principled definition for identifying ridges of high probability density. For a mixture of three Gaussians in two dimensions, the principal curves and the subspace-gradient ascent trajectories are illustrated in Figure 1. We provide a summary comparison of the original local maxima identification conditions and the local principal manifold identification conditions in Table 1. These are the two local necessary and sufficient conditions for a point to belong to a principal manifold of specified dimension. For the following let $p(\mathbf{x})$ be the continuous and twice differentiable probability density function of the random vector of interest, $\mathbf{g}(\mathbf{x})$ its gradient-transpose, and $\mathbf{H}(\mathbf{x})$ its Hessian matrix evaluated at a

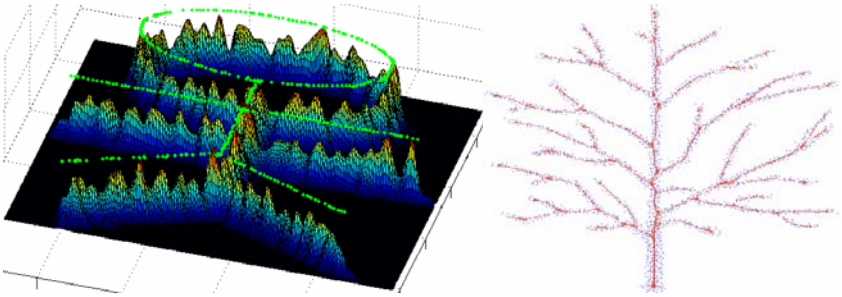

Fig. 2. The 1-dimensional principal manifolds of the diffused stickman distribution (green in left picture) and that of a diffused tree distribution (red in right picture), both identified using subspace mean-shift iterations.

particular point **x**. Also let $\{\lambda_i(\mathbf{x}),\mathbf{q}_i(\mathbf{x})\}$, $i=1,\ldots,n$ be the eigendecomposition of the Hessian (the dependence of these on the evaluation point **x** will not be explicitly denoted in the following if clear from the context).

Definition 2.1. [13] A point **x** is an element of the d-dimensional principal set, denoted by P^d iff $\mathbf{g}(\mathbf{x})$ is orthogonal (null inner product) to at least $(n-d)$ eigenvectors of $\mathbf{H}(\mathbf{x})$ and $p(\mathbf{x})$ is a strict local maximum in the subspace spanned by these $(n-d)$ eigenvectors (eigenvalues corresponding to these eigenvectors are strictly less than zero).

This definition states explicitly the conditions in Table 1 and lead to interesting properties regarding principal manifolds, such as the nonlinear deflation property ($P^d \subset P^{d+1}$) and local maxima being appointed as the 0-dimensional principal manifold. Consequently, principal curves pass through local maxima, principal 2-dimensional surfaces pass through principal curves, etc. Another natural consequence of this definition is the simple criterion for checking whether a point is on the principal curve or not. Specifically, the principal curves are characterized by points at which the gradient becomes an eigenvector of the Hessian (i.e., $\mathbf{Hg}=\lambda\mathbf{g}$) and all the other eigenvalues of **H** are negative. For iterative hill-climbing algorithms such as subspace gradient ascent or subspace mean-shift, this identity could be utilized to form a suitable stopping criterion to detect when the trajectory is in the vicinity of the principal curve.

The definition also highlights the potential complications that one might encounter in a general NBSS problem, such as self-intersecting, bifurcating, or looping principal curves. These occurrences are illustrated on two datasets in Figure 2, and are generally avoided by researchers addressing nonlinear coordinate unfolding problems. At this time, the only feasible approach to unify the local coordinate systems formed by each segment of a principal curve seems to utilize an atlas structure, stitching piecewise coordinate systems at boundaries. Limiting our discussion to simpler situations here the principal manifolds also form a global nonlinear coordinate frame at least in the domain defined by the bounded support of the source distribution, we

can employ nonlinear manifold unfolding techniques and utilize geodesic or line-of-curvature based differential geometric measures of metrics to define an isometric nonlinear transformation of the mixture data into an Euclidean coordinate frame, in a manner similar to Lee et al [5] and Harmeling et al [6].

Specifically focusing on linear mixtures, it is straightforward to verify that the principal lines of a prewhitened mixture of independent sources with unimodal zero-mean densities coincide with the linear ICA solution. To see this one can check that after prewhitening only rotation remains, which does not change the geometric properties of the joint density of the sources, thus the structural principal lines defined via subspace maximum likelihood remains unchanged except for a coordinate rotation. For the special case of a jointly Gaussian density, this also means that the proposed subspace maximum likelihood nonlinear principal components coincide with the usual linear principal components.

3 Nonlinear Coordinate Unfolding for NBSS

We demonstrate the proposed nonlinear coordinate unfolding technique that is applicable to any distributions, but most conveniently employed for distributions that are symmetrically and unimodally distributed around a simple manifold structure that unwrap in a single piece to a global Euclidean frame. While the technique applies to general twice differentiable densities, we will illustrate a specific implementation for a Gaussian mixture model.

Consider a Gaussian data distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. The logarithm of this density expressed as a quadratic form in $\boldsymbol{\delta}$, perturbation from a point $\mathbf{x}$, is obtained easily with some algebraic manipulation as

$$\log G(\mathbf{x}+\boldsymbol{\delta};\boldsymbol{\mu},\boldsymbol{\Sigma}) = [\gamma_0 + \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{x} - \mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x}/2] + [\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1} - \mathbf{x}^T\boldsymbol{\Sigma}^{-1}]\boldsymbol{\delta} - \boldsymbol{\delta}\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}/2 \quad (2)$$

for any perturbation $\boldsymbol{\delta}$ from $\mathbf{x}$. Since the PCA projections of a data point from a joint Gaussian to a lower dimensional principal (linear) manifold follow trajectories along the eigenvectors of $\boldsymbol{\Sigma}^{-1}$, we seek to create an analogy with this for projecting points drawn from arbitrary distributions to their corresponding nonlinear principal manifolds. For an arbitrary pdf $p(\mathbf{x})$, with gradient $\mathbf{g}$ and Hessian $\mathbf{H}$ at the point of interest (along a projection trajectory), we observe from Taylor's expension up to the quadratic term that

$$\log p(\mathbf{x}+\boldsymbol{\delta}) \approx \log p + (\mathbf{g}/p)^T\boldsymbol{\delta} + \boldsymbol{\delta}^T[\mathbf{H}/p - (\mathbf{g}/p)(\mathbf{g}/p)^T]\boldsymbol{\delta}/2 \quad (3)$$

where p, $\mathbf{g}$, $\mathbf{H}$ are all evaluated at $\mathbf{x}$. Equating terms in (2) and (3), we obtain that the local mean and the local covariance inverse of $p(\mathbf{x})$ is given by

$$\begin{aligned} \boldsymbol{\Sigma}^{-1}(\mathbf{x}) &\approx -\mathbf{H}(\mathbf{x})/p(\mathbf{x}) + (\mathbf{g}(\mathbf{x})/p(\mathbf{x}))(\mathbf{g}(\mathbf{x})/p(\mathbf{x}))^T] \\ \boldsymbol{\mu}(\mathbf{x}) &\approx \mathbf{x} + \boldsymbol{\Sigma}(\mathbf{x})\mathbf{g}(\mathbf{x})/p(\mathbf{x}) \end{aligned} \quad (4)$$

The GMM illustration in Fig. 1 is instructive to understand how one might use the principal curves as a means of measuring curvilinear local orthogonal coordinates. For simplicity of discussion let's focus on the 2-dimensional data case here. Specifically, at each local maximum, the principal curves form a locally Euclidean orthogonal coordinate frame (green). Starting from an arbitrary point $\mathbf{x}$, one can trace

out subspace gradient ascent trajectories (blue) to project this point to the corresponding principal point. The subspace gradient ascent simply follows the eigenvector direction at the initial point $\mathbf{x}$ to which the projection of the gradient at $\mathbf{x}$ is maximum. The eigenvector of choice is the eigenvector of $\boldsymbol{\Sigma}^{-1}$ based on the discussion regarding analogies with Gaussian densities and linear PCA projections. The projection trajectory is simply traced out by solving the following differential equation

$$\begin{aligned} &\dot{\mathbf{y}}(t) = \mathbf{q}_m(\mathbf{y}(t))\mathbf{q}_m^T(\mathbf{y}(t))\mathbf{g}(\mathbf{y}(t)) \\ &\text{where } m = \arg\min_i \left\| \dot{\mathbf{q}}_i(\mathbf{y}(t)) \right\|, \quad \mathbf{y}(0) = \mathbf{x} \\ &\text{and } \mathbf{q}_m(\mathbf{y}(0)) = \arg\max_{\{\mathbf{q}_i(\mathbf{x})\}_{i=1,2}} |\mathbf{g}^T(\mathbf{x})\mathbf{q}_i(\mathbf{x})| \end{aligned} \tag{5}$$

Until the condition $\mathbf{Hg}=\lambda\mathbf{g}$ is satisfied (which is equivalently stated as $\boldsymbol{\Sigma}^{-1}\mathbf{g}=\gamma\mathbf{g}$, since $\mathbf{g}$ is both an eigenvector of $\mathbf{H}$ and $\mathbf{gg}^T$). This differential equation solves for the trajectory initialized to $\mathbf{x}$ and tangent to the eigenvector $\boldsymbol{\Sigma}^{-1}(\mathbf{x})$ that points towards the direction of maximal rate of increase of $p(\mathbf{x})$ among all orthogonal directions given by the candidate eigenvectors. The trajectory converges to a point $\mathbf{x}_p$ on the principal curve.[1] The length of the curvilinear trajectory from $\mathbf{x}$ to $\mathbf{x}_p$ can be appointed as the coordinate of $\mathbf{x}$ in the direction orthogonal to the principal curve. Taking an arbitrary reference point on the principal curve (which can now be traced by solving a differential equation that follows the eigenvector that is parallel to the gradient) as the origin (e.g., the local maximum in the middle of the three in Fig. 1), one could also measure the distance of $\mathbf{x}_p$ along the principal curve to this origin, yielding the second coordinate of $\mathbf{x}$ with respect to the global coordinate frame formed by the nonlinear principal curve. Note that these geometrically simple global curvilinear coordinates are only possible in a very limited set of scenarios and researchers have typically dealt with these simplified cases due to lack of an understanding of how to systematically globalize many piecewise local orthogonal curvilinear coordinate frames in challenging scenarios such as the stickman or the tree examples in Fig. 2.

Shortest Path Along the Principal Curve: For a given finite dataset drawn from a known or estimated density $p(\mathbf{x})$, once all data points are projected onto the principal curve using (5), the projections form a smooth one-dimensional manifold that could be approximated by a sparse connected graph such as a minimum spanning tree, k-nearest-neighbor graph, or an ε-ball graph. The *geodesic* distance between any two points on the graph could be determined with a shortest path algorithm, such as Dijkstra's greedy search [14]. A tempting idea is to employ sparse connected graph based approximations for approximately finding the projection lengths from $\mathbf{x}$ to $\mathbf{x}_p$, however this idea would not work on a graph formed by the original data points since not enough samples might lie sufficiently close to the sought principal curve. A possibility could be to iterate a small but sufficiently large number of additional

[1] Convergence proof is relatively trivial. The pdf always increases since the derivative always points in a direction that has positive inner product with the gradient and when the trajectory reaches a point on the principal curve, the gradient becomes the eigenvector of $\boldsymbol{\Sigma}^{-1}(\mathbf{y}(t))$ itself (but the one that is orthogonal to the trajectory), thus the stopping criterion is achieved.

points $\mathbf{z}$ using (5) to obtain $\mathbf{z}_p$ as a roughly uniform sample from the principal curve and include these in graph construction (i.e. union of $\mathbf{x}$ and $\mathbf{z}_p$), such that the distance from every data point $\mathbf{x}$ to every principal curve sample $\mathbf{z}_p$ can be approximated with a fast shortest path search method and te one that is closest (in a suitable sense) can be appointed as its projection. The graph formed only using the set $\mathbf{z}_p$ could then be utilized to find the second coordinate.

Gaussian Mixture Models for Nonlinear Coordinate Unfolding: Suppose that a set of independent and identically distributed (iid) samples are available: $\{\mathbf{x}_1,\ldots,\mathbf{x}_N\}$. Assume that a Gaussian mixture model (GMM) fit in the form

$$p(\mathbf{x}) = \sum_m \alpha_m G(\mathbf{x}; \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \tag{6}$$

is obtained using established density estimation techniques with proper attention paid to model order selection. The gradient and Hessian of this pdf estimate has a convenient self-similar form that facilitates algorithm design. Specifically, we have

$$\begin{aligned} \mathbf{g}(\mathbf{x}) &= -\sum_m \alpha_m G(\mathbf{x}; \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \boldsymbol{\Sigma}_m^{-1} (\mathbf{x} - \boldsymbol{\mu}_m) \\ \mathbf{H}(\mathbf{x}) &= \sum_m \alpha_m G(\mathbf{x}; \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) [\boldsymbol{\Sigma}_m^{-1} (\mathbf{x} - \boldsymbol{\mu}_m)(\mathbf{x} - \boldsymbol{\mu}_m)^T \boldsymbol{\Sigma}_m^{-1} - \boldsymbol{\Sigma}_m^{-1}] \end{aligned} \tag{7}$$

which leads to the expression

$$\boldsymbol{\Sigma}^{-1}(\mathbf{x}) = \sum_k w_k(\mathbf{x}) \boldsymbol{\Sigma}_k^{-1} + \sum_k \sum_{l \neq k} w_k(\mathbf{x}) w_l(\mathbf{x}) \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)(\mathbf{x} - \boldsymbol{\mu}_l)^T \boldsymbol{\Sigma}_l^{-1} \tag{8}$$

for the local covariance, where $\mathbf{w}_k(\mathbf{x}) = \alpha_k G(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) / p(\mathbf{x})$. Clearly, for points close to the mean of a particular Gaussian component with respect to the Mahalanobis distance involving the corresponding component covariance, the second term on the right hand side of (8) becomes negligible with respect to the first term, thus the local nonlinear coordinate frame approaches an orthogonal Euclidean frame centered at the local maximum, as expected. This observation provides a theoretical motivation for local PCA, which will also clearly fail in transient regions between components according to this theory.

Case Study with Periodic Sources: We demonstrate the nonlinear unfolding strategy on a case study that uses a 20-component GMM to approximate the density of a mixture that is obtained by spiral-wrapping of two sources with respective sinusoid and piecewise linear sources that are periodic with relatively prime frequencies (7Hz and 13Hz) respectively (see Fig. 3). Specifically, following example 5.3 of [6] (because it cannot be reduced to a post- or pre-nonlinear mixture problem) we had $s_1(t)=0.5\sin(14\pi t)$, $s_2(t)=\arctan(\sin(26\pi t)/\cos(26\pi t))/\pi$ and $z=6s_1+s_2+6$, $x_1=z\cos(3\pi s_1)$, $x_2=z\sin(3\pi s_1)$. Time index t is sampled in the interval [-0.5,0.5] at 1000Hz. The unfolding algorithm described above is employed. The mixture samples are projected onto their corresponding coordinate points along the principal curve using the differential equation in (5) using the GMM model as the basis for principal curve estimation. Once the principal curve projections are obtained, Dijkstra's algorithm is applied to the 1-ball neighborhood graph consisting of Euclidean distance between pairs connected according to the graph in order to get the relative coordinates along the principal curve. The results of this case study are presented in Fig. 3. Apart from some convergence-related noise and the expected nonlinear distortion of the source signals that cannot be recovered without source distribution information, the proposed technique is reasonably successful in identifying the original source signals.

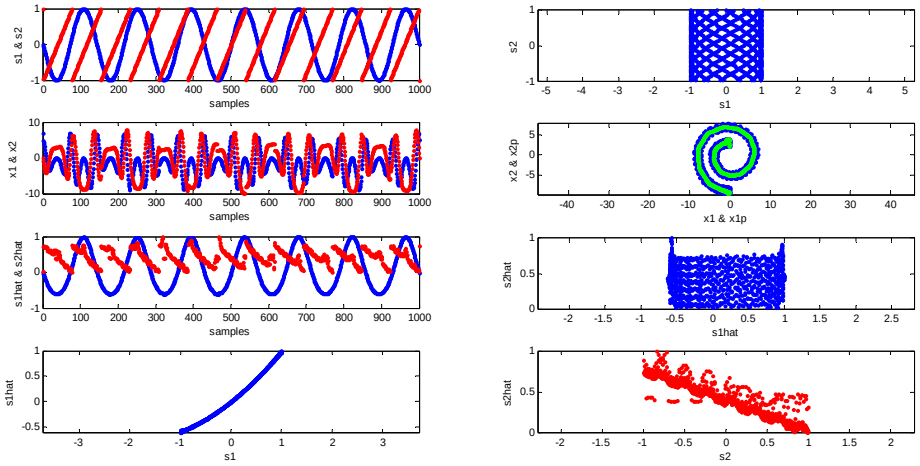

Fig. 3. Summary of coordinate unfolding results for the periodic sources. Rows from left to right: (1) Source signals versus time and the source distribution; (2) mixtures versus time and mixtures with respect to each other; (3) Unfolded sources versus time and the unfolded source distribution; (4) Estimated source 1 versus true source 1 and estimated source 2 versus true source 2 (ideally a monotonic curve is desired after correcting for permutation).

However, we note that this case study involves a nice global spiral principal curve that enables us to determine a global Euclidean unfolding solution that would be impossible otherwise.

Case Study with Random Sources: We present results with the same nonlinear mixture as in the previous case study, but replace the periodic (approximately) orthogonal sources with independent random sources that have Uniform (support [-1,1]) and Gaussian (0 mean, 0.2 standard deviation) distributions. The number of samples in this illustration is 1000. The procedure for identifying the unfolded coordinates is identical with the previous case study and the results are summarized in Fig. 4.

4 Conclusions

Nonlinear blind source separation is a challenging problem that has not yet been formulated satisfactorily to yield a unique well-defined solution. The literature on nonlinear independent components primarily focuses on the relatively trivial extension of linear ICA, referred to as postnonlinear mixture separation. More recent attempts to utilize ideas from manifold learning (for instance, isometric dimensionality reduction and kernel principal component analysis), have not clearly discussed the challenges involved in finding the intricate details of algorithms that

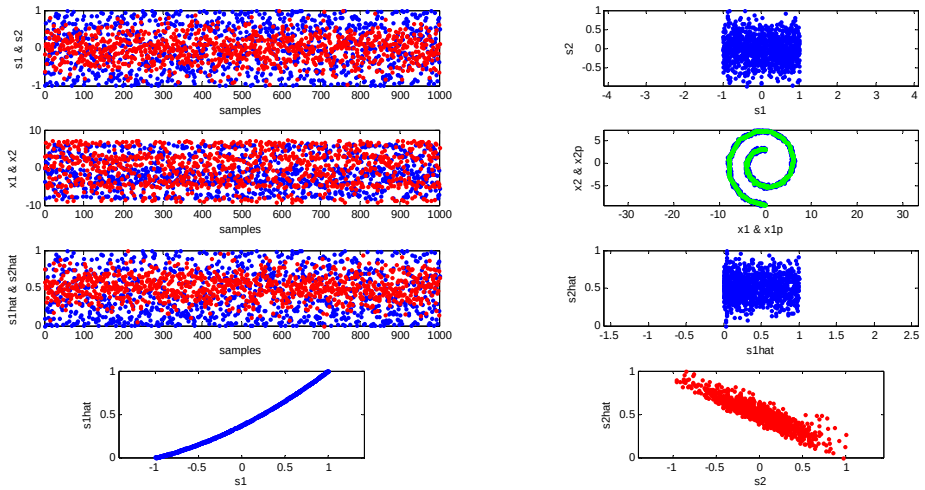

Fig. 4. Summary of coordinate unfolding results for the periodic sources. Rows from left to right: (1) Source signals versus time and the source distribution; (2) mixtures versus time and mixtures with respect to each other; (3) Unfolded sources versus time and the unfolded source distribution; (4) Estimated source 1 versus true source 1 and estimated source 2 versus true source 2 (ideally a monotonic curve is desired after correcting for permutation).

will work in various scenarios – in fact there are many scenarios where nonlinear coordinate unfolding as proposed in such papers will not generalize to outside the limited set of geometries they consider. In this paper we aimed to achieve two goals: (i) point out some unmentioned caveats in nonlinear blind source separation using manifold learning (ii) present the application of a maximum likelihood type principal curve identification technique to the problem of coordinate unfolding in a differential geometric framework. Results obtained using a nonlinearity mixture used by researchers in another paper have shown that the unfolding technique is promising, the proposed principal curve coordinate system can recover sources under the assumption of unimodal variations around a global (in the support of source densities) curvilinear manifold.

Acknowledgments. This work is partially supported by NSF grants ECS-0524835, ECS-0622239, and IIS-0713690.

References

1. Hyvarinen, A., Pajunen, P.: Nonlinear Independent Component Analysis: Existence and Uniqueness Results. Neural Networks. 12(3) (1999) 429-439
2. Almeida, L.: MISEP – Linear and Nonlinear ICA based on Mutual Information. Journal of MachineLearning Research. 4 (2003) 1297-1318

3. Parra, L., Deco, G., Miesbach, S.: Statistical Independence and Novelty Detection with Information Preserving Nonlinear Maps. Neural Computation. 8 (1996) 260-269
4. Jutten, C., Karhunen, J.: Advances in Blind Source Separation (BSS) and Independent Component Analysis (ICA) for Nonlinear Mixtures. Int. J. Neural Systems. 14(5) (2004) 267-292
5. Lee, J.A., Jutten, C., Verleysen, M.: Nonlinear ICA by Using Isometric Dimensionality Reduction. Proceedings of ICA 2004, LNCS 3195 (2004) 710-717
6. Harmeling, S., Ziehe, A., Kawanabe, M., Muller, K.R.: Kernel Based Nonlinear Blind Source Separation. Neural Computation. 15 (2003) 1089-1124
7. Hastie, T., Stuetzle, W.: Principal Curves. Jour. Am. Statistical Assoc., 84 (406) (1989) 502-516
8. Tibshirani, R.: Principal Curves Revisited. Statistics and Computation. 2(1992) 183-190
9. Sandilya, S., Kulkarni, S.R.: Principal Curves with Bounded Turn. IEEE Trans. on Information Theory. 48(10) (2002) 2789-2793
10. Kegl, B., Kryzak, A., Linder, T., Zeger, K.: Learning and Design of Principal Curves. IEEE Trans. on PAMI. 22(3) (2000) 281-297
11. Stanford, D.C., Raftery, A.E.: Finding Curvilinear Features in Spatial Point Patterns: Principal Curve Clustering with Noise. IEEE Trans. on PAMI. 22(6) (2000) 601-609
12. Chang, K., Grosh, J.: A Unified Model for Probabilistic Principal Surfaces. IEEE Trans. on PAMI. 24(1) (2002) 59-74
13. Erdogmus, D., Ozertem, U.: Self-Consistent Locally Defined Principal Curves. Proceedings of ICASSP 2007. 2 (2007) 549-552.
14. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithms. 2nd edn. MIT Press and McGraw-Hill (2001)