Anisotropic diffusion on sub-manifolds with application to Earth structure classification

Dan Kushnir *, Ali Haddad, Ronald R. Coifman

Dept. of Mathematics, Yale university, 51 Prospect st., New Haven, CT 06511, USA

ARTICLE INFO	ABSTRACT

We introduce a method to re-parameterize massive high dimensional data, generated by nonlinear mixing, into its independent physical parameters. Our method enables the identification of the original parameters and their extension to new observations without any knowledge of the true physical model. The suggested approach in this paper is related to spectral independent components analysis (ICA) via the construction of an anisotropic diffusion kernel whose eigenfunctions comprise the independent components. However, we use a novel anisotropic diffusion process, utilizing only a small observed subset $\bar{Y}$, that approximates the isotropic diffusion on the parametric manifold $\mathcal{M}_X$ of the full set Y. We employ a Nyström-type extension of the independent components of $\bar{Y}$ to the independent components of Y, and provide a validation scheme for our algorithm parameters choice. We demonstrate our method on synthetic examples and on real application examples.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

In many applications one would like to recover the underlying independent parameters that give rise to high dimensional observed data. Given such data, many embedding techniques have been suggested to reduce the data complexity and represent it in a lower dimension (see [2]). However, in many situations a method that embeds the data into its independent components may be the method of choice, since it guarantees to represent the data in the unique space of its independent physical parameters. Independent components analysis (ICA) [1] provides such a fundamental tool. The linear ICA problem is formulated mathematically as follows. Let the sources $x^1, x^2, \ldots, x^d$ be d unknown independent random variables, and A be a $D \times d$ unknown constant mixing matrix, such that $D \gg d$. The task is to find the mixing matrix, and thus also the sources, from N different observations of the D-vector y: $y = Ax$. In nonlinear ICA instead of the linear matrix A, one observes some nonlinear mixing transformation: $y = f(x)$. The ICA, as a re-parametrization method, is a useful tool for solving empirically inverse problems. In particular when the underlying physical model is unknown. In learning tasks (such as semi-supervised classification), ICA provides a meaningful space to compute relations between data points. In this paper we provide an efficient algorithm for computing an extendable independent re-parameterization of high dimensional data.

In order to solve the nonlinear ICA problem, certain assumptions on the generation of data need to be met. Similarly to [4], we assume that the data is generated by independent Itô processes in the parametric space, and then transformed to a higher dimensional observed space by some nonlinear transformation $f : X \to Y$. f is an arbitrary nonlinear transformation, however, we assume that it is smooth and bi-Lipschitz: there exists a constant $K > 0$ such that

$$\frac{1}{K} \left\| x^{(i)} - x^{(j)} \right\| \leqslant \left\| f\left(x^{(i)}\right) - f\left(x^{(j)}\right) \right\| \leqslant K \left\| x^{(i)} - x^{(j)} \right\| \tag{1}$$

---

* Corresponding author.

E-mail addresses: dan.kushnir@yale.edu (D. Kushnir), ali.haddad@yale.edu (A. Haddad), coifman-ronald@yale.edu (R.R. Coifman).

for all $x^{(i)}, x^{(j)} \in X$. Thus, both f and its inverse are differentiable almost everywhere. The smoothness allows one to use local clouds generated in the parametric space and mapped to the observed space, to estimate the local distortion of the parametric manifold, by using the differential J_f. The differential-based distortion is estimated by the local covariance matrix, which is used within a metric we construct to estimate the Euclidean distances between points in the parametric space. The local metric enables the construction of a diffusion kernel that approximates the isotropic diffusion on the parametric data manifold $\mathcal{M}_X$, from which the normalized Laplacian L of the data graph is constructed. L asymptotically converges to the a separable Fokker–Planck (FP) operator $\mathcal{L}$ on the parametric manifold. A subset of the eigenfunctions of $\mathcal{L}$ are monotonic functions of the independent variables, as guaranteed by Stürm–Liouville oscillation theorem [3]. Thus, the eigenfunctions form independent components (ICs) and can be used for data re-parametrization.

In the context of supervised learning, the task of extending functions on the data to new observations often arises. A variety of extension techniques via eigenfunctions of data driven matrices have been employed and investigated e.g. [8–12]. In this paper we address the specific problem of extending the independent components representation from a challenging small *reference set* $\bar{Y}$ to a larger set containing newly observed data – Y. To this end, our work contributes a generalization of the nonlinear ICA [4] in the sense that it computes the ICs of a-priory given data, but also provides an efficient extension for the ICs to any newly observed data. We emphasize that the metric used in [4] within the diffusion kernel relies on estimates of the local distortions of the parameters in the observable space. However, such distortions may not be available for all observations. In our work we extend [4], and propose an efficient extendable spectral ICA algorithm. In our algorithm we use a different intrinsic metric between the observations, which depends only on estimates of the local distortions of just a few reference points. The ability to extend the embedding enables an efficient recovery of the independent parameters of observations which are not available in advance. For this purpose we construct an anisotropic diffusion kernel – W on $\bar{Y}$ that is a Hilbert–Schmidt operator, differently than the construction in [4]. W is compact on $\bar{Y}$, self-adjoint, and has a discrete sequence of eigenvalues and eigenvectors that can be extended by a Nyström-type extension to new points in Y. The extension itself yields an orthogonal set of eigenfunctions. These eigenfunctions asymptotically approximate the independent components of Y.

The mapping from the observable space into the independent components space can be seen as an inverse problem, in which the observed data is transformed back to the input parameters of f. Given empirical values of some function ('labeling') on $\bar{Y}$, an interpolation or extension of that function to new points can be done based on data affinities computed on the linearized manifold in the IC-space, whereas data affinities computed on the nonlinear manifold in the observable space may yield erroneous results. We thus exploit interpolation in IC-space to compute f for new points on the parametric manifold by merely interpolating known function values. This procedure is also used to validate the re-parametrization.

We demonstrate our method on a synthetic example as well as on classification of electro-magnetic (EM) measurements associated with layered geological models. In the last application we also demonstrate that our ICA extension allows significant sub-sampling of the data manifold while maintaining high accuracy of classification. In particular, we show that the classification done in the IC-space is better than the one obtained by classifying with a Euclidean metric in the observable space.

The mapping into independent components also suggests an invariance with respect to different sensors measuring the same physical phenomena. In other words, we observe the same mapping into the parameters space even when the measurements are done by different sensors. The invariance can be used as a mean for efficient selection of sensors, or, when it does not exist, as a mean to define a regime of the independent parameters for which a particular subset of sensors is sensitive.

The paper is organized as follows. In Section 2 we describe the nonlinear ICA problem and its setting. In Section 3 the construction of anisotropic diffusion kernels on subsets and their role in ICA is described. Section 4 entails the extension scheme of ICA to new observations and includes a synthetic example. Also, a validation method for the extension parameters is suggested. In Section 5 we describe the application of our method for the classification of EM measurements associated with layered geological models.

2. ICA of Itô processes

In [3] and [4] Singer and Singer and Coifman describe a spectral approach for solving linear and nonlinear ICA problems, respectively. In [4] the authors assume that the data is generated by stochastic Itô processes in order to obtain a unique solution to the generally ill-posed nonlinear ICA problem. This assumption enables the computation of the local Jacobian-based distortion metric induced by the nonlinear transformation that maps the parameter space into the observable space. As shown in [4], a spectral approach for solving the nonlinear ICA problem can then be employed, using the Jacobian based metric, to construct a diffusion kernel whose eigenvectors are the independent components. In this section we briefly describe a similar setting for data generation by Itô processes.

Problem definition. At the setting of our problem is an unknown parametric manifold $\mathcal{M} \subseteq X \subset \mathbb{R}^d$ and a corresponding observed data set Y. Y is generated by a nonlinear bi-Lipschitz map $f : X \rightarrow Y$, such that Y is embedded in a high dimension $Y \subset \mathbb{R}^D$. The dimension d of the data parametric manifold is lower or equal to the dimension of the observable space ($d \leqslant D$). The variables x^i are independent Itô processes, given by

$$dx^i = a^i(x^i)dt + b^i(x^i)dw^i, \quad i = 1, \ldots, d, \tag{2}$$

where a^i and b^i are drift and noise coefficients and w^i is Brownian motion. The unknown nonlinear map $y^j = f^j(x^1, \ldots, x^d)$, $j = 1, \ldots, D$ (often called the mixing transformation), is to be found together with the processes x^i. Moreover, we would like to extend the inverse map of f for new observations $y = f(x)$ after it was already computed for a small set.

Local covariance and the Jacobian. We define the $D \times D$ covariance matrix of an observed process to be

$$C = JB^2J^T, \tag{3}$$

where B is the $d \times d$ diagonal noise matrix with $B_{ii} = b^i$, and J is the Jacobian matrix $J_{ij} = f_i^j = \frac{\partial f^j}{\partial x^i}$. The matrix B can be assumed to be the identity by applying a change of variables such that

$$d\tilde{x}^i = \tilde{a}^i(x^i)dt + 1dw^i. \tag{4}$$

This is equivalent to rescaling the mixing transformation to assume the sources have unit variance. Using Itô's lemma we can write the covariance matrix as

$$C = JJ^T. \tag{5}$$

Clearly, the Jacobian of f is not accessible. All is available for us is the observed data and the local covariance matrices.

3. Anisotropic diffusion on sub-manifolds

In this section we describe the construction of an anisotropic diffusion kernel on the observed manifold of $\bar{Y} \subseteq Y$ that approximates the isotropic diffusion kernel on the parametric manifold $\bar{X}$. The diffusion kernel we construct has the desired attribute that it is separable, and its first (non-trivial) eigenfunctions are monotonic functions of the independent parameters. The construction of the discrete operator is done for a minimal sub-sample of the manifold, later shown to have a straightforward extension to Y.

3.1. Approximating Euclidean distances on the parametric manifold

The Euclidean distances between two observed points in X can be approximated by using the Jacobian of the nonlinear map. We take a different strategy than in [4] and use the Jacobian at the mid-point to estimate the Euclidean distance.

Let $x, \xi \in X$ be two points in the parametric space X, and $f : X \to Y$ a nonlinear map such that $y = f(x)$ and $\eta = f(\xi)$. Define $g : Y \to X$ to be the inverse map of $f : X \to Y$, g can be approximated by a Taylor series at the point $\frac{\eta+y}{2}$ (i.e. $g(\frac{\eta+y}{2}) = \delta$):

$$\xi^i = \delta^i + \frac{1}{2}\sum_j g_j^i\left(\frac{\eta+y}{2}\right)(\eta^j - y^j) + \frac{1}{8}\sum_{kl} g_{kl}^i\left(\frac{\eta+y}{2}\right)(\eta^k - y^k)(\eta^l - y^l) + O(\|\eta - y\|^3),$$

where $g_j^i = \frac{\partial g^i}{\partial y^j}$. Considering a similar expansion for x^i, the squared Euclidean distance in parametric space can be approximated by

$$\|\xi - x\|^2 = \sum_i (\xi^i - x^i)^2 = \sum_{ijk} g_j^i\left(\frac{\eta+y}{2}\right)g_k^i\left(\frac{\eta+y}{2}\right)(\eta^j - y^j)(\eta^k - y^k) + O(\|\eta - y\|^4). \tag{6}$$

In matrix notation the second-order approximation to the Euclidean distance is

$$\|\xi - x\|^2 = (\eta - y)^T\left[(JJ^T)^{-1}\left(\frac{\eta+y}{2}\right)\right](\eta - y) + O(\|\eta - y\|^4). \tag{7}$$

The following Lemma 3.1 provides a second-order approximation to the Jacobian at the mid-point:

Lemma 3.1. *Let J be the Jacobian of the bi-Lipschitz function $f : X \to Y$, and let $x, y \in X$ such that $y = f(x)$, $\eta = f(\xi)$. Then*

$$(JJ^T)^{-1}\left(\frac{\eta+y}{2}\right) = 2[JJ^T(\eta) + JJ^T(y)]^{-1} + O(\|\eta - y\|^2). \tag{8}$$

Proof is given in Appendix A.

Using Lemma 3.1 in Eq. (7) yields the second-order approximation to the Euclidean distance on X

$$\|\xi - x\|^2 = 2(\eta - y)^T [JJ^T(\eta) + JJ^T(y)]^{-1}(\eta - y) + O(\|\eta - y\|^4). \tag{9}$$

We note that in [4] a Taylor expansions around y and η are given, which yield the second-order approximation

$$\|\xi - x\|^2 = \frac{1}{2}(\eta - y)^T [(JJ^T)^{-1}(\eta) + (JJ^T)^{-1}(y)](\eta - y) + O(\|\eta - y\|^4). \tag{10}$$

The expansion at the mid-point yields a different approximation – (9), which allows the construction of a different kernel than in [4]. The new kernel provides *a straightforward extension* of the spectral ICA to newly observed data points. A property which does not exist for the kernel suggested in [4].

3.2. The diffusion operator on the sub-manifold and its limiting operator

Construction of the integral operator. Consider an m-sample of *reference points* $\bar{Y} = \bar{y}^{(1)}, \ldots, \bar{y}^{(m)}$ from the observed data $y^{(1)}, \ldots, y^{(N)} \in Y \subseteq \mathbb{R}^D$, where $m \ll N$. Typically, $\bar{Y}$ are chosen so that $\bar{X}$ are known. As we show below, $\bar{X}$ can then be extended to the rest of Y. For now, we only assume that $\bar{Y}$ and Y are generated from $\bar{X}$ and X respectively, by the same nonlinear transformation f, as described in Section 2. We compute the $N \times m$ affinity matrix between the sample and the set of size N points

$$A_{ij} = \exp\left(-\frac{\|J^{-1}(\bar{y}^{(j)})(\bar{y}^{(j)} - y^{(i)})\|^2}{\varepsilon}\right), \quad i = 1, \ldots, N, \ j = 1, \ldots, m. \tag{11}$$

A is computed using the equality

$$\|J^{-1}(\bar{y}^{(j)})(\bar{y}^{(j)} - y^{(i)})\|^2 = (\bar{y}^{(j)} - y^{(i)})^T (JJ^T)^{-1}(\bar{y}^{(j)})(\bar{y}^{(j)} - y^{(i)}). \tag{12}$$

We compute the $m \times m$ matrix

$$W = \omega^{-\frac{1}{2}} A^T A \omega^{-\frac{1}{2}} \tag{13}$$

that comprises the anisotropic diffusion processes between the reference set $\bar{Y}$ via the points in Y. The density normalization used in (13) is applied by using the $m \times m$ diagonal matrix with the sums of $A^T A$'s column entries: $\omega = \text{diag}(A^T A \mathbf{1})$, on its diagonal, where $\mathbf{1}$ is an $m \times 1$ column vector with all entries equal to 1. The normalization corresponds to the approximation of the FP operator and its eigenfunctions via W [14], that is pursued further below.

The following result approximates the matrix (13) in the continuous limit by a kernel employing a distortion-based metric with the inverse of the sum of Jacobians at $\bar{y}^{(i)}$ and $\bar{y}^{(j)}$:

Theorem 3.2. *The kernel*

$$W_{ij} = \int\limits_Y \exp\left\{-\frac{\|J^{-1}(\bar{y}^{(i)})(\bar{y}^{(i)} - y)\|^2 + \|J^{-1}(\bar{y}^{(j)})(\bar{y}^{(j)} - y)\|^2}{\varepsilon}\right\} dy \tag{14}$$

corresponding to the matrix (13) *can be approximated to a second order by*

$$W_{ij} = \frac{\sqrt{\pi}}{\sqrt{\det(J^T(\tilde{y})J(\tilde{y}))}} \exp\left\{-\frac{(\bar{y}^{(j)} - \bar{y}^{(i)})[JJ^T(\bar{y}^{(i)}) + JJ^T(\bar{y}^{(j)})]^{-1}(\bar{y}^{(j)} - \bar{y}^{(i)})}{\varepsilon}\right\}, \tag{15}$$

where $\tilde{y} = \frac{\bar{y}^{(i)} + \bar{y}^{(j)}}{2}$, *and* $\bar{y}^{(i)}, \bar{y}^{(j)} \in \bar{Y}$.

Proof is given in Appendix B.

Employing Theorem 3.2, it is straightforward to show that as $m \to \infty$ the discrete operator (13) converges to the integral operator

$$\frac{1}{m} \sum_{j=1}^m W_{ij} q_{\bar{Y}}(\bar{y}^{(j)}) \to \int\limits_{\bar{Y}} \exp\left\{\frac{(y - \bar{y}^{(i)})[JJ^T(\bar{y}^{(i)}) + JJ^T(y)]^{-1}(y - \bar{y}^{(i)})}{\varepsilon}\right\} p_{\bar{Y}}(y) q_{\bar{Y}}(y) \, dy, \tag{16}$$

where $q_{\bar{Y}} : \bar{Y} \to \mathbb{R}$ is some function, $p_{\bar{Y}}(y)$ is the density, the constant π is ignored, and we used the fact that $\frac{1}{\sqrt{\det(J(\tilde{y})J(\tilde{y})^T)}}$ approximates the density at $\tilde{y}$. The 2nd order approximation suggested in Eq. (9) can be used now so that (16) is equal to the integral on the parametric space $\bar{X}$

$$\int_{\bar{X}} \exp\left\{ -\frac{\|\bar{x}^{(i)} - x\|^2 + O(\|\bar{x}^{(i)} - x\|^4)}{2\varepsilon} \right\} p_{\bar{X}}(x) q_{\bar{X}}(x)\, dx, \tag{17}$$

where $p_{\bar{X}}(x)$ is the density in $\bar{X}$.

Convergence to the backward Fokker–Planck operator and spectral ICA. The normalized graph Laplacian

$$L = D^{-1}W - I \tag{18}$$

is constructed from W (16), with $D = \mathrm{diag}(W\mathbf{1})$, where $\mathbf{1}$ is an $m \times 1$ column vector with all entries equal to 1. $\frac{1}{\varepsilon}L$ converges to the backward FP operator $\mathcal{L}$ on $\bar{X}$ [4,14], with Neumann boundary condition:

$$\mathcal{L}q = \triangle q - \nabla U \cdot \nabla q, \tag{19}$$

where the potential U depends on the density $U = -2\log p_{\bar{X}}$. Since the density in $\bar{X}$ is a product of the one-dimensional densities $p^i(\bar{x}^i)$, the potential satisfies $U(\bar{x}) = \sum_i U^i(\bar{x}^i)$. Therefore $\mathcal{L}$ separates into n one-dimensional operators:

$$\mathcal{L} = \triangle + 2\nabla \log p_{\bar{X}} \nabla = \sum_i \left(\frac{\partial^2}{\partial^2 \bar{x}^i} + 2\frac{\partial \log p^i(\bar{x}^i)}{\partial \bar{x}^i} \frac{\partial}{\partial \bar{x}^i} \right) = \sum_i \mathcal{L}_i, \tag{20}$$

where $\mathcal{L}_i$ corresponds to the 1-dimensional backward FP operator on the data. A set of eigenfunctions $\Phi_d = [\varphi_1, \ldots, \varphi_d]$ that correspond to the first non-trivial eigenfunctions of the 1D operators $\mathcal{L}_i$ in (20) are monotonic functions of the processes x^i, as guarantied by Stürm–Liouville oscillation theory [3]. In the discrete case these are the eigenvectors of L that provide a *re-parameterization* of the data in terms of its independent parameters. In other words, these eigenfunctions are the *independent components*. The map Φ_d can also be interpreted as the *inverse map* of our nonlinear transformation f, up to a local scaling.

We note that the d eigenfunctions are not necessarily the first d non-trivial eigenvectors corresponding to the d minimal eigenvalues of L. To detect the pure eigenvectors we can use the method suggested in [3], or use the pairwise mutual information between the eigenvectors.

4. An extendable ICA algorithm

In this section we show that a robust and efficient extension of the independent components on $\bar{Y}$ to new observables in Y can be obtained by using the singular value decomposition (SVD) of A. We start with discussing the SVD in the setting of reproducing kernel Hilbert space [7]. The SVD provides the algebraic relation between the normalized matrices A (Eq. (11)) and W (Eq. (13)): the eigenvectors of W are the right singular vectors of A. The extension of the eigenvectors of W for new observables is then done via a Nyström-type method. The extended eigenfunctions form an orthogonal set that approximately preserves the statistical independence property. Thus, it forms an ICA for the points outside the reference set. We demonstrate the extension on a synthetic example and show that a straightforward validation of the results and parameter tuning can be performed as well.

4.1. The restriction and extension operators

For deriving the extension of the independent components we first establish a connection between the operator construction in Section 3.2 to reproducing kernels Hilbert space [7]. We consider again the m-sample $\bar{Y} = \{\bar{y}^{(i)}\}_{i=1}^m$ of $Y = \{y^{(i)}\}_{i=1}^N$ such that $\bar{Y} \subset Y \subset \mathbb{R}^D$, and let μ be a measure on Y. Let $\hat{W}$ be a symmetric, semi-positive definite, bounded kernel $\hat{W}: Y \times Y \to \mathbb{R}$. Following [7] there exists a unique reproducing kernel Hilbert space $\mathcal{H}$ of functions defined on Y for which $\hat{W}$ is a reproducing kernel. To facilitate notation we denote $\tilde{A} = \omega^{-\frac{1}{2}} A^T$. Then the density-normalized operator $\tilde{A}: L^2(\bar{Y}, \omega^{-\frac{1}{2}}d\mu) \to \mathcal{H}$ (known as the extension operator) and its adjoint $\tilde{A}^*: \mathcal{H} \to L^2(\bar{Y}, \omega^{-\frac{1}{2}}d\mu)$ (restriction operator) can be used to construct the Hilbert–Schmidt operator $\tilde{A}^*\tilde{A}: L^2(\bar{Y}, \omega^{-1}d\mu) \to L^2(\bar{Y}, \omega^{-1}d\mu)$, defined in (13). The operator $\tilde{A}^*\tilde{A}$ is compact, self-adjoint, and has a discrete sequence of eigenvalues and eigenvectors.

To draw the algebraic connection between $\tilde{A}\tilde{A}^*$ and $\tilde{A}^*\tilde{A}$ we revisit the singular value decomposition: the SVD of the $N \times m$ matrix $\tilde{A}$ produces a singular system of strictly positive singular values λ_i, $i = 1, \ldots, k$, where $k = rank(\tilde{A})$, and vectors $\{\varphi_j\}_{j=1}^m \in \mathbb{C}^m$ and $\{\psi_j\}_{j=1}^N \in \mathbb{C}^N$ that form an orthonormal basis of $\mathbb{C}^m$ and $\mathbb{C}^N$, respectively. In matrix form $\tilde{A} = \Psi \Lambda \Phi^*$ such that

$$\tilde{A}\varphi_j = \sqrt{\lambda_j}\psi_j, \quad j = 1, \ldots, k,$$
$$\tilde{A}\varphi_j = 0, \quad j = k+1, \ldots, m,$$
$$\tilde{A}^*\psi_j = \sqrt{\lambda_j}\varphi_j, \quad j = 1, \ldots, k,$$
$$\tilde{A}^*\psi_j = 0, \quad j = k+1, \ldots, N,$$

and

$$\tilde{A}^*\tilde{A}\varphi_j = \lambda_j\varphi_j, \quad j = 1,\dots,k,$$
$$\tilde{A}^*\tilde{A}\varphi_j = 0, \quad j = k+1,\dots,m,$$
$$\tilde{A}\tilde{A}^*\psi_j = \lambda_j\psi_j, \quad j = 1,\dots,k,$$
$$\tilde{A}\tilde{A}^*\psi_j = 0, \quad j = k+1,\dots,N.$$

The utility of $\tilde{A}^*$ is underlined in the following lemma:

Lemma 4.1. *Let $F \in \mathcal{H}$, then $\tilde{A}^*F = f$, where $f = F|_{\bar{Y}}$.*

The proof of Lemma 4.1 is based on the reproducibility property [7] and given in various papers such as [9]. In our context we would like to find an extension of f to F by minimizing

$$\inf_F \left\| f - \tilde{A}^*F \right\|^2_{L^2(\bar{Y},\omega^{-\frac{1}{2}}d\mu)} + \gamma\|F\|^2_{\mathcal{H}}, \tag{21}$$

where $\gamma > 0$ is a regularization parameter.

The following observations are motivated by the SVD formulation and (21):

1. If $f = \sum_k a_k\varphi_k$, then the minimizer of (21) is given by

$$F_\gamma = \sum_k \frac{\lambda_k}{\lambda_k + \gamma} a_k\psi_k. \tag{22}$$

The error is

$$\left\| f - \tilde{A}^*F \right\|_{L^2(\bar{Y},\omega^{-\frac{1}{2}}d\mu)} = \sum_k \frac{\gamma^2}{(\lambda_k + \gamma)^2} a_k^2. \tag{23}$$

2. $\tilde{A}^*\tilde{A}$ and $\tilde{A}\tilde{A}^*$ are positive, self-adjoint matrices. The spectra of $\tilde{A}^*\tilde{A}$ and $\tilde{A}\tilde{A}^*$ are the same, and as $\gamma \to 0$ we recover the Nyström extension:

$$\psi_j(y) = \frac{1}{\sqrt{\lambda_j}} \int_{\bar{Y}} A(y,s)\varphi_j(s)\,d\mu(s), \quad y \in Y. \tag{24}$$

3. From Lemma 4.1 and (24), the eigenfunctions of $\tilde{A}^*\tilde{A}$ and $\tilde{A}\tilde{A}^*$ coincide on $\bar{Y}$.
4. ψ is the extension of φ outside the set $\bar{Y}$.
5. The extension preserves the orthonormality of the eigenfunctions (with respect to the measure $\omega^{-\frac{1}{2}}d\mu$).

We conclude that the extension Ψ_d of the independent components Φ_d on $\bar{Y}$ yields independent components on Y since each ψ_i, $i = 1,\dots,d$, is a weighted combination of the independent φ_i, such that as $\varepsilon \to 0$ each ψ_i converges to φ_i. Finally, we emphasize the computational gain obtained by computing Ψ via (24) instead of using a direct method for the eigen-decomposition of the matrix $\tilde{A}\tilde{A}^*$.

4.2. Extension error

Given the true parameters of $\bar{Y}$ (namely, $\bar{X}$), the re-parametrization error of a point $y^{(i)} \in Y$ is computed as the difference between its interpolated parameters values and its true parameters $x^{(i)}$. Specifically, the parameters of $y^{(i)}$ are approximated by

$$\tilde{x}^{(i)}_{new} = \sum_{k:\Psi_d(\bar{y}^{(k)})\in N_i} c^{(i,k)}\bar{x}^{(k)}, \tag{25}$$

where N_i are the k-nearest neighbors of $\Psi_d(y^{(i)})$ computed with the Euclidean metric on Ψ_d, and $c^{(i,k)}$ are the linear interpolation coefficients computed in Ψ_d. The error of the re-parametrization is

$$err_{prm}\big(\Psi_d\big(y^{(i)}\big)\big) = \big\|x^{(i)} - \tilde{x}^{(i)}_{new}\big\|^2. \tag{26}$$

The interpolation error when using an interpolation in the observables space Y is also considered.

$$\hat{x}_{new}^{(i)} = \sum_{k:\bar{y}^{(k)} \in N_i} c^{(i,k)} \bar{x}^{(k)}, \tag{27}$$

where $c^{(i,k)}$ are linear interpolation coefficients computed in Y. The corresponding error is measured by

$$err_{obs}(y^{(i)}) = \|x^{(i)} - \hat{x}_{new}^{(i)}\|^2. \tag{28}$$

4.3. Synthetic example

We demonstrate the spectral ICA and its eigenvector extension with a synthetic example. The example employs the nonlinear map from polar coordinates to cartesian coordinates in the 2D plane. We generate 500 reference points in the parameters space $\bar{X} = \{(r^{(i)}, \theta^{(i)})\}_{i=1}^{500}$ such that $r \sim U[1, 2]$, and $\theta \sim \mathcal{N}(0, 1)$ is conditioned to take values only within $[-\frac{\pi}{4}, \frac{\pi}{4}]$.

For each point $\bar{x}^{(i)} = (r^{(i)}, \theta^{(i)})$, we generate a local Gaussian cloud such that each point $x_c^{(i)}(r, \theta)$ in the cloud is generated by

$$r_c^{(i)} = r^{(i)} + \sqrt{dt}\, w^{(i)},$$
$$\theta_c^{(i)} = \theta^{(i)} - \frac{\theta^{(i)}}{2} dt + \sqrt{dt}\, w^{(i)}, \tag{29}$$

with $dt = 0.1$. The reference points and their corresponding clouds are then mapped by the nonlinear mapping $f(x) = y$

$$\bar{y}^{(i)} = \left(r^{(i)} \cos(\theta^{(i)}), r^{(i)} \sin(\theta^{(i)})\right) \tag{30}$$

and

$$y_c^{(i)} = \left(r_c^{(i)} \cos(\theta_c^{(i)}), r_c^{(i)} \sin(\theta_c^{(i)})\right) \tag{31}$$

to its cartesian coordinates. The clouds are then used to compute the local covariance matrix in the observable space $\bar{Y}$. 2000 additional points are generated from a similar distribution for which the independent components extension is demonstrated below. The points in the parametric space and in the observed space are plotted in Fig. 1A, B, respectively.

Next, we construct the weight matrix W (Eq. (13)), using the local sample covariance matrix to approximate JJ^T via Eq. (5), and $\varepsilon = 0.2$. The right eigenvectors $\{\varphi_i\}_{i=1}^2$ of the Markov matrix $D^{-1}W$ are then computed. These eigenvectors approximate the eigenvectors of the FP operator, as discussed in Section 3.2. The first eigen-pair (φ_0, λ_0) corresponding to the eigenvalue $\lambda_0 = 1$ is a constant $\varphi_0 = [1, 1, \ldots, 1]$ and thus disregarded. The next two eigenvectors are monotonic functions of r and θ [3]. We construct the embedding of $\bar{Y}$ using $\Phi_2 : \bar{y}^{(i)} \to [\varphi_1(\bar{y}^{(i)}), \varphi_2(\bar{y}^{(i)})]$. That is, every point $\bar{y}^{(i)}$ is mapped to a vector in $\mathbb{R}^2$ containing the i-th coordinate of the first two (non-trivial) eigenvectors. In Fig. 1C, D the embedding of the reference points is illustrated, where the color coding corresponds to the true parameters r and θ. The color coding suggests that the embedding coordinates comprise the independent components of the set $\bar{Y}$. In sub-figures E and F we plot the extended points (only), where the color coding demonstrates that the left eigenvectors ψ_1 and ψ_2 are independent components of the set Y.

4.4. Validation via synthesis of new observations

Extension of functions on the data can be done in various ways (see for example [9–11]). We consider the extension of the coordinates in the observable space $\bar{Y}$ to a new point $x \in X$, namely, synthesizing a new observable. For the sake of simplicity, the extension of a function on X can be done by interpolation within a patch around x. The interpolation reflects the linear weights between points on $\mathcal{M}_X$ and can be used to synthesize a new observable $y \approx f(x)$ by

$$\Psi_d^{-1}(x) = \sum_{k:\bar{x}^{(k)} \in N_x} c_k \bar{y}^{(k)}, \tag{32}$$

where $\bar{x}^{(k)}$ are neighbors of x, $\bar{y}^{(k)}$ are their corresponding coordinates in Y, and c_k are the interpolation coefficients. In our setting the parametric space X is inaccessible to us. However, we can choose a point $\tilde{x}$ in the re-parametrization Ψ_d and synthesize a corresponding observable by using the interpolation formula in (32) in Ψ_d instead of X. The interpolation weights computed in the eigen-space of Ψ_d are expected to be similar to the interpolation weights in the true parametric space X.

Eq. (32) can be used as a tool to validate the accuracy of our inverse map into the eigen-space of independent components. The validation is done in the following way: given a point $y \in Y$ and its map $\Psi_d = [\psi_1(y), \ldots, \psi_d(y)]$, we can validate the accuracy of our re-parametrization by checking whether an observable mapped to the eigen-space Ψ_d is mapped back to its original coordinates y. In practice, we can use the validation to find optimal values of the algorithm parameters by minimizing the error

$$err = \|y - \Psi_d^{-1}(x)\|^2. \tag{33}$$
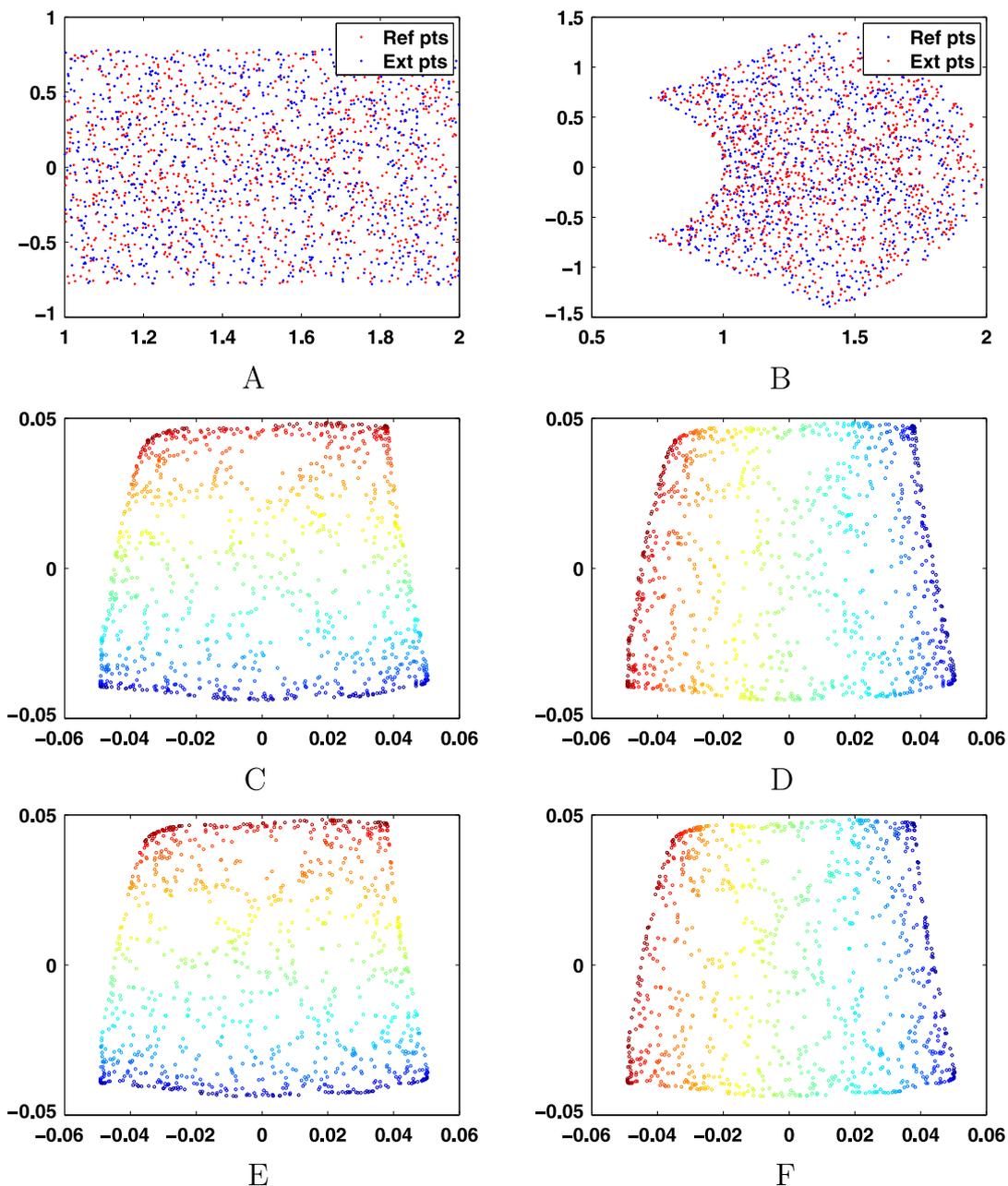
Fig. 1. A synthetic example. A, B: 2D illustration of data in parametric space (A), and in observed space (B). C, D: embedding of reference points into Φ_2 with color coding corresponding to r (C), and θ (D). E, F: Embedding of the extended points into Ψ_2 with color coding corresponding to r (E), and θ (F).

We demonstrate the optimization of the parameter ε for the synthetic example in Section 4.3. The interpolation coefficients c_k in (32) are computed by

$$c_k(x) = \frac{\exp\left(-\frac{\|x - \bar{x}^{(k)}\|^2}{\sigma_{\Psi_d(y)}}\right)}{\sum_{i : \bar{x}^{(i)} \in N_{\bar{X}}} \exp\left(-\frac{\|x - \bar{x}^{(i)}\|^2}{\sigma_{\Psi_d(y)}}\right)}. \tag{34}$$

The choice of weights in (34) is motivated by the isotropic diffusion process defined on $\bar{X}$, as shown in Eq. (17). The mean error (33) is computed for 100 randomly sampled points $y^{(i)}$ in the synthetic example described in Section 4.3. $\sigma_{\Psi_d(y^{(i)})}$ is chosen to be the minimal distance between $\Psi_d(y^{(i)})$ and its nearest neighbor. The parameter σ can also be added to the optimization process, however, for simplicity we keep it fixed and minimize the error (33) with respect to ε only. The minimal error $err = 0.05$ is attained at $\varepsilon = 0.2$. The error is of order $O(\varepsilon^2)$ as implied from the approximation (9).

288

D. Kushnir et al. / Appl. Comput. Harmon. Anal. 32 (2012) 280–294

4.5. Coping with measurement noise

In the physical setting of our nonlinear ICA problem additive Gaussian measurement noise may be considered. Since the measurement tool is controlled by the user, the measurements noise can be estimated and calibrated to be of a normal distribution. It is straightforward to show, via the Fourier transform, that the noisy matrix $\tilde{W}$ computed via (13) is a Gaussian kernel with the covariance

$$\tilde{C} = \left[J J^T \left(y^{(i)} \right) + J J^T \left(y^{(j)} \right) + \sigma I \right]. \tag{35}$$

We can then subtract σI from the measured covariances to compute the filtered matrix W and recover the independent components.

Clearly, the correct estimation of the noise covariance matrix from given data is limited and depends on the size of the data sample (see, for example, [15,16]). However, as mentioned above, we assume that the user has control on the tool and that enough measurements can be acquired for a known set of parameters, so that the noise variance can be estimated to a sufficient accuracy.

4.6. The algorithm

The complete description of our algorithm is summarized in Algorithm 1.

Algorithm 1 Re-parametrization algorithm

Training stage:

1. Obtain N measurements Y of data.
2. Extract m measurements $\bar{Y}$ corresponding to (known) reference samples of the independent parameters $\bar{X}$, and their covariance matrices $C(y^{(i)})$.
3. Compute the affinity matrix $\mathbf{W}$ according to (13), for a given kernel scale ε.
4. Employ eigenvalue decomposition of $\mathbf{D}^{-1}\mathbf{W}$ (18) and obtain the eigenvalues $\{\lambda_j\}$ and the eigenvectors $\{\boldsymbol{\varphi}_j\}$.
5. Construct the map $\boldsymbol{\Phi}_d$ to obtain a re-parametrization of the reference observations.
6. Construct the inverse map $\boldsymbol{\Phi}_d^{-1}$ according to (32).
7. Find the optimal kernel scale ε that minimizes (33), by repeating 3–6 for different scales.

Testing stage:

1. Given a set of new observations $y^{(i)}$, compute the normalized affinity matrix $\tilde{\mathbf{A}}$.
2. Extend $\boldsymbol{\psi}_j$ as a weighted combination of $\boldsymbol{\varphi}_j$ via (24).
3. Construct the map $\boldsymbol{\Psi}_d$ to obtain a re-parametrization of the new observations.
4. Recover the independent parameters according to (25) and compute the mean re-parametrization error (26).

---

5. Classification of layered Earth model formations

In this section we demonstrate our ICA-based re-parametrization and its extension on simulated directional electro–magnetic (EM) measurements that have been used in oilfield applications, an introduction to which can be found in [13]. The signals are measured while traversing geological layers using a tool consisting of antennas and receivers. The EM measurements are sensitive to the position of Earth layer boundaries relative to the measurement device as well as to the resistivities of the corresponding Earth layers. The electromagnetic measurements are suitable for demonstrating our method for the following reasons:

1. while the number of individual measurement channels can be quite large (10–30), the number of geological parameters describing the Earth model is typically much smaller (1–6);
2. this type of measurements suggests the existence of local clouds in parametric space which allows us to invert the nonlinear transformation and obtain the independent geological parameters. The clouds emerge since the composition of layers is typically not homogeneous, which gives rise to local perturbation in the resistivity parameters. Also, boundaries between layers are typically not smooth and various scales of irregularity may exist. Such irregularities give rise to perturbations in the distance between the measurement tool and the boundary.

For the examples shown in this paper, we consider a set of simulated measurements generated using a two layer Earth formation model illustrated in Fig. 2, where R_i is the i-th layer resistivity and h is the distance from the tool to the boundary with the lower layer. For simplicity, we assume that the horizontal resistivity of the layer containing the tool is known (pH = 6.3 Ohm m) and that all formation layers are isotropic, that is, their horizontal and vertical resistivities are equal. The measurements themselves can be modeled as pairs of tilted transmitter and receiver dipoles that operate at various frequencies (100 and 400 kHz are typical frequencies) and spacings (96 and 34 inches are typical coil spacings). In
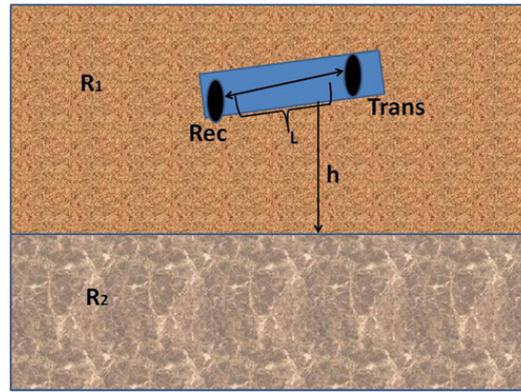
Fig. 2. Two-layers parametric model. Layer's resistivity is denoted by R, whereas the distance between the tool layer and the boundary is denoted by h.

addition, the EM signals are measured using complex voltage ratios, so that an individual measurement channel is either a phase (in units of degrees) or an amplitude (in units of decibels (db)).

5.1. Re-parametrization of EM measurements

An open problem in the interpretation of the EM measurements, is the direct construction of the mapping between electromagnetic measurements and the corresponding geological parameters. One approach that could be used is based on a lookup table: a set of measurements $\bar{Y}$ has its own labeling comprised of the true parameters, and a new observable y is assigned with labeling by using interpolation from the closest table entries. The use of a lookup table is based on employing the Euclidean metric in the observable space. When the measurements exhibit nonlinearity, the Euclidean distances do not reflect any informative proximity between the data points, unless the manifold is over-sampled. As shown below, use of Euclidean metric in Y leads to erroneous results, when the task is to re-parameterize observed data, or when supervised classification/parameterization of new observables is performed.

ICA and extension. We demonstrate the performance of our extension on a data set of two-layer geological models in which the tool is traveling near an interface with another layer. The governing parameters are the distance h of the tool to the other layer and the layer resistivity R_2. We use two tool channels: one with spacing of $96''$ and 100 kHz, and the other with spacing $34''$ and 400 kHz. We generated a set of 961 reference points $\bar{Y} = \{\bar{y}^{(i)}\}_{i=1}^{961} \in \mathbb{R}^4$ such that $\bar{y}^{(i)}$ is a vector

$$\bar{y}^{(i)} = (Att_1, PS_1, Att_2, PS_2)^T \tag{36}$$

of measurements taken from the two channels that corresponds to the model parameters

$$\bar{x}^{(i)} = (R_2, h, R_1)^T \tag{37}$$

where R_1 is held fixed for all i at 6.3095. The parameters h and R_2 are sampled with equidistant spacing within the rectangle $[0.2, 10]_{ft} \times [0.2, 6.3095]_{\Omega m}$. Thus, the number of independent sources in this problem is 2. Next, a sample $\{y^{(j)}\}_{j=1}^{24025}$ of observables is generated such that each 25 observed points are generated from a Gaussian distribution around each reference point.

We demonstrate below the extension of the data re-parametrization. First, each $\bar{y}^{(i)}$ corresponds to a Gaussian cloud from which the local covariance matrix $C(y^{(i)})$ is computed. We can thus approximate $J(y^{(i)})J^T(y^{(i)})$ using (3) for the construction of the affinity matrix A between the reference points $\bar{Y}$ and Y (see Eq. (11)). We next construct the 961×961 matrix W (Eq. (13)). The $d = 3$ right eigenvectors $[\varphi_0, \varphi_1, \varphi_2]$ of $D^{-1}W$ are computed, corresponding to the eigenvectors of the normalized graph Laplacian

$$L = D^{-1}W - I. \tag{38}$$

The first eigen-pair is disregarded, whereas the next two eigenvectors approximate the eigenvectors of the FP operator and thus are monotonic functions of h and R_2, i.e. the independent components. We construct the re-parametrization of $\bar{Y}$ using $\Phi_2 : \bar{y}^{(i)} \to [\varphi_1(\bar{y}^{(i)}), \varphi_2(\bar{y}^{(i)})]$: $\bar{y}^{(i)}$ is mapped to a vector in $\mathbb{R}^2$ containing the i-th coordinate of the first two (non-trivial) eigenvectors of (38). Next, we construct the extension to Y, $\Psi_2 : y^{(i)} \to [\psi_1(y^{(i)}), \psi_2(y^{(i)})]$, by using the extension formula (24).

We plot the reference points $\bar{X}$ in the 2D parametric space in Fig. 3A, and the complete data set parameters – X in Fig. 3B. The x-axis corresponds to R_2 and the y-axis corresponds to h. To demonstrate the nonlinearity of f, the observed reference points are plotted in Fig. 3C, D. In Fig. 3E and F we plot the reference points re-parametrization Φ_2 and its extension to Ψ_2, respectively.
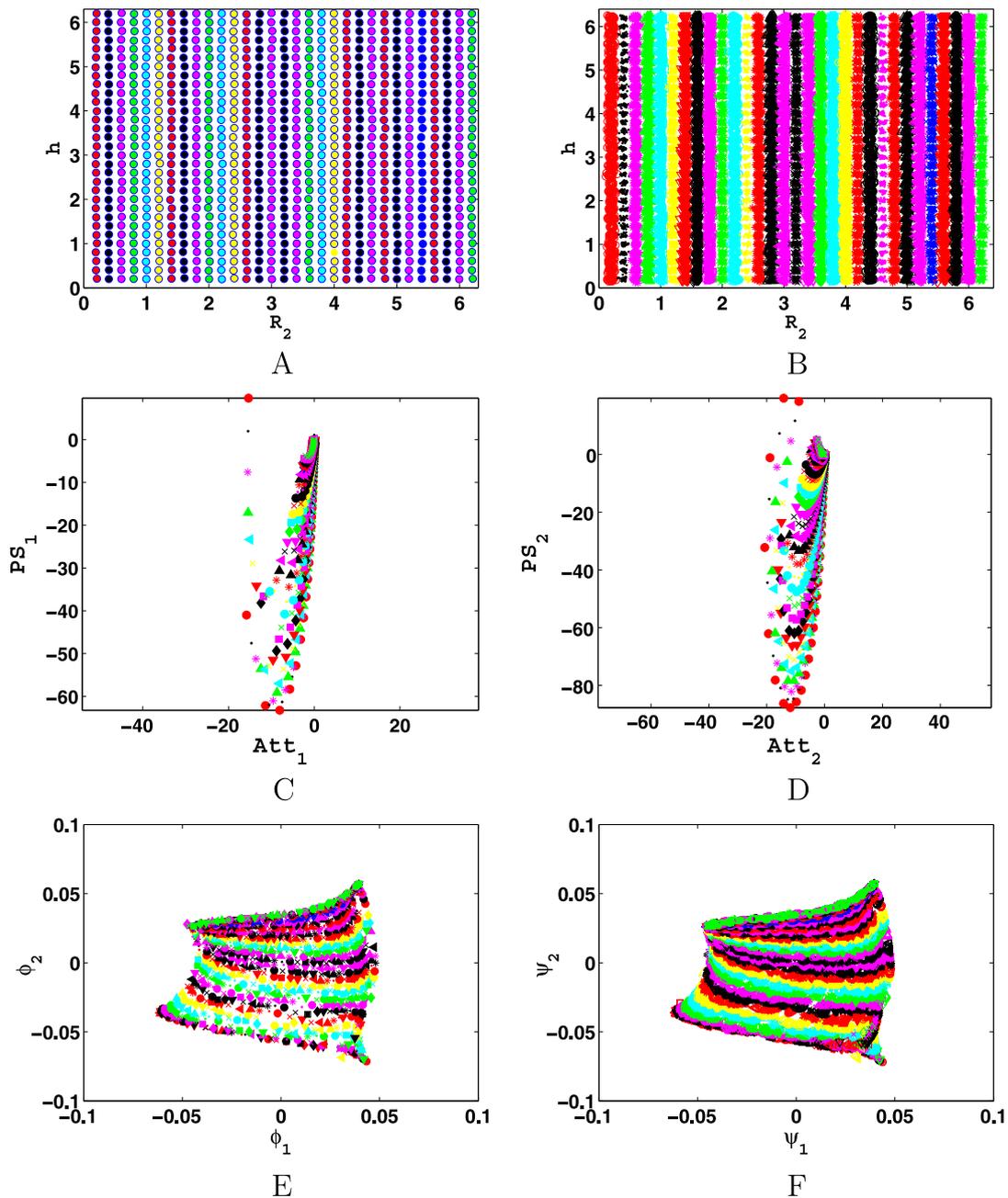
Fig. 3. Embedding and extension of EM data. A, B: The 2D parametric space of h and R_2 sampled. $\bar{X}$ (A), X (B). C, D: Observed data $\bar{Y}$. 2D illustration of observed data in 96″ channel (C). 34″ channel (D). E, F: Data re-parametrization. Embedding of reference points $\bar{Y}$ by Φ_2 (E). The extension Ψ_2 (F).

Extension error. Given the true parameters of $\bar{Y}$ (namely, $\bar{X}$), the re-parametrization error of a point $y^{(i)} \in Y$ is computed as the difference between its interpolated parameters values and its true parameters $x^{(i)}$.

In the following we show that the number of reference points can be significantly reduced while sustaining a low re-parametrization error – $err(y^{(i)})$ (26). Whereas the interpolation error using Euclidean metric in the observable space (28) manifests high errors. Also, we show that the ICA-based re-parametrization error is comparable to an error that is based on fitting of a Gaussian Mixture Model (GMM) to the reference points (and their associated covariances). In this last case the interpolation coefficients in (26) are based on the same distortion based metric used in (7). For the purpose of this demonstration, $\bar{Y}$ is sub-sampled uniformly for fewer and fewer points. The remaining set $Y \smallsetminus \bar{Y}$ is used as a pool of points for testing our extension. In Fig. 4 the error is plotted against the rate of sampling of the two parameters R_2 and h. In Fig. 5 we present histograms of the errors for a sampling rate of $r = 0.2$, demonstrating that except for a few errors, the classification error employing our ICA extension is very low, where a deeper examination reveals that the few high magnitude errors take place only on the boundaries of the data set, where our embedding suffers from some artifacts. The error based on GMM interpolation yields similar errors since it considers the local distortion around each reference points, however, this method in general cannot reproduce a coordinate system in the independent parameters. On the other hand, the classification error using a Euclidean metric in Y is almost uniformly distributed on the interval $[0, 2]$.
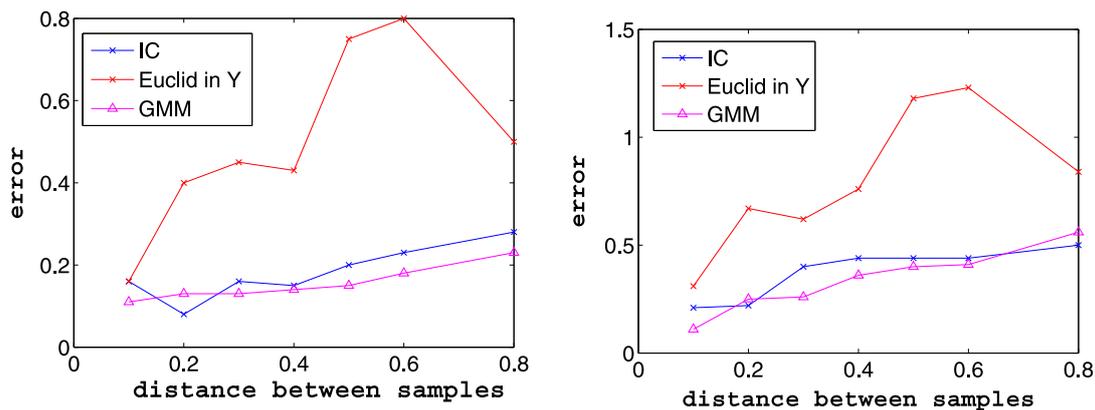
Fig. 4. Extension errors of ICA-based re-parametrization, a Euclidean-based lookup table, and a GMM-based re-parametrization. Left: mean errors of resistivity – R_2. Right: mean errors of distance to boundary – h.
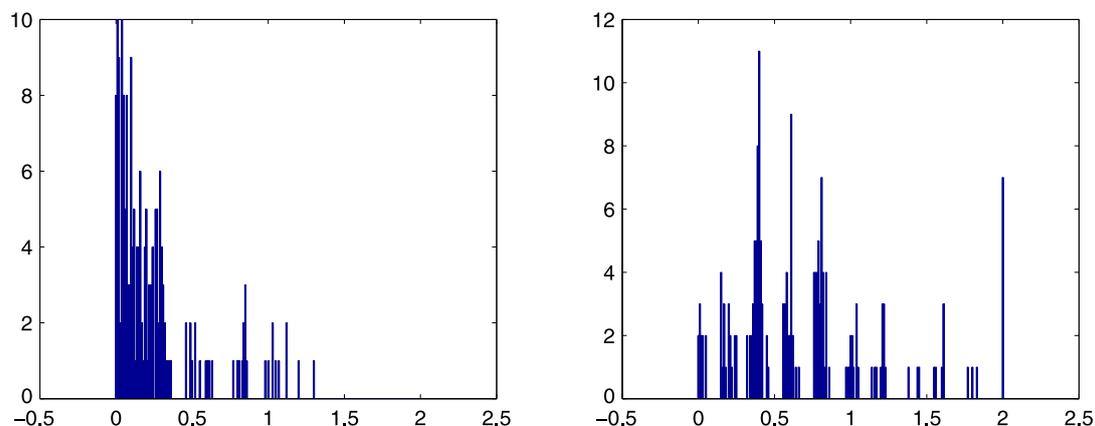
Fig. 5. Histograms of extension error of ICA-based re-parametrization and a Euclidean-based lookup table. Left: errors of h for our method. Right: error of h for the Euclidean metric in Y.

Feature space selection. We demonstrate that our ICA method can be used as a tool for the selection of the feature space – the channels being used for optimal re-parametrization and classification. For this matter, we generate the first data set using two channels of 96 inches with 100 and 400 kHz. The second data set is generated using one channel of 96 inches with 100 kHz and another channel of 34 inches with 400 kHz. We also generate a data set in 6D involving the union of all channels. The mappings of the two 4D data sets into the independent components are very similar to each other. While different channels exhibit different sensitivity to certain regimes of parameters, the majority of the data is mapped similarly into the two components, as shown in Fig. 6, where a close view of the interior of the manifold is shown. Thus manifesting the invariancy of the sensors choice with respect to the true physical parameters. The union of the six channels demonstrates an embedding that has less distortion then the distortion in the embedding obtained from subsets of channels, as seen in Fig. 7.

Observations. The results reported above motivate the following observations:

- The observed data is nonlinear with varying density.
- The re-parametrization is an ICA. Namely, the x-axis corresponds to variations in the h coordinate only, whereas the y-axis corresponds to variations in the R_2 coordinate only.
- The re-parametrization approximates the rectangular domain of the original parameters.
- Interpolation or classification of new observables is more accurate via the re-parametrization then in the observables space.
- As the sampling rate increases, err_{prm} is lower than the re-parametrization error in the observable space (27) – err_{obs}. Thus suggesting an efficient compression of any lookup table by using our extendable ICA.
- The histograms reveal that for most points the error err_{prm} is very low, and very few high errors are located near the boundaries of the data set.
- The mapping into independent components reveals the invariancy to the measurements, namely, although the same physical phenomena is measured by different sensors, the mapping into the independent components is similar.
- Using more channels may compensate for subsets of channels which yield distorted parameterizations. On the other hand, adding unsensitive channels may also deteriorate good parameterizations. Thus, the ICA-based re-parametrization
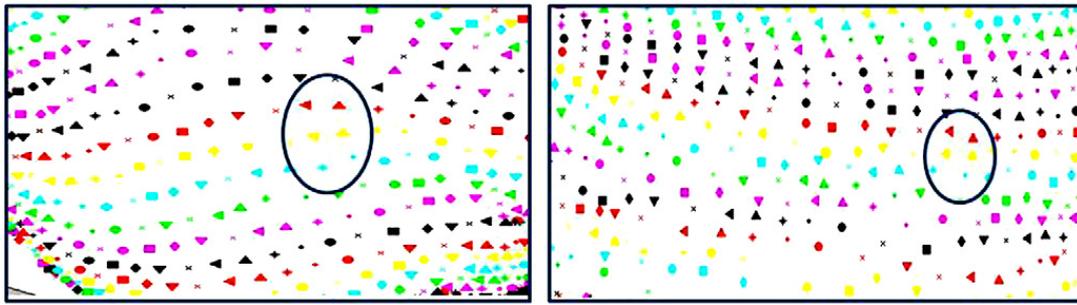
Fig. 6. Similar maps of different channel sets into their independent components. Left: embedding of channels of distance 96 inches with frequencies 100 and 400 kHz. Right: embedding of channels of distance 96 and 34 inches with frequencies 100 and 400 kHz, respectively. A sampled region of correspondence is circled in both embedding.
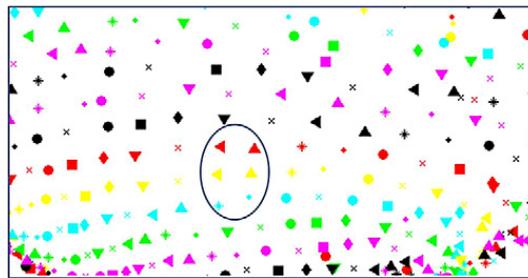
Fig. 7. Similar maps of different channel sets into their independent components: embedding of all channels. A sampled region of correspondence is circled.

suggests a tool for choosing a feature space that yields satisfactory ICA parameterizations, or identifying regimes of parameters where a particular selection of features does not.

6. Conclusion

We have described a method for the computation and extension of independent components of high dimensional data generated by a nonlinear transformation of independent variables. Our assumption is that the variables in the parametric space are governed by Itô processes that give rise to local clouds. The clouds enable the approximation of the FP operator on the parametric manifold by the graph Laplacian. The eigenfunctions of the resulting graph Laplacian are independent components for which a Nyström-type extension can be used to generate an ICA for the rest of the space. The results obtained suggest that our method can be used for semi-supervised learning, in particular because it enables to construct efficiently a physically meaningful coordinate system for new observations. We demonstrated the advantage of our method for the classification of EM measurements of Earths layers structure, showing better classification results then the results of classification performed in the observable space.

Acknowledgment

Appendix A. Proof for Lemma 3.1

Proof. Expanding $C(x) = JJ^T(x)$ and $C(\xi) = JJ^T(\xi)$ in a Taylor series near the point $\frac{x+\xi}{2}$ yields

$$C(x) = C\left(\frac{x+\xi}{2}\right) + \nabla C\left(\frac{x+\xi}{2}\right)\left(\frac{x-\xi}{2}\right) + \frac{1}{2}\nabla^2 C\left(\frac{x+\xi}{2}\right)\left(\frac{x-\xi}{2}\right)^2 + O\left(\left(\frac{x-\xi}{2}\right)^3\right), \tag{A.1}$$

$$C(\xi) = C\left(\frac{x+\xi}{2}\right) + \nabla C\left(\frac{x+\xi}{2}\right)\left(\frac{\xi-x}{2}\right) + \frac{1}{2}\nabla^2 C\left(\frac{x+\xi}{2}\right)\left(\frac{\xi-x}{2}\right)^2 + O\left(\left(\frac{\xi-x}{2}\right)^3\right). \tag{A.2}$$

Adding the two equations yields

$$C(x) + C(y) = 2C\left(\frac{x+\xi}{2}\right) + O\left(\left(\frac{\xi-x}{2}\right)^2\right). \tag{A.3}$$

D. Kushnir et al. / Appl. Comput. Harmon. Anal. 32 (2012) 280–294

293

To facilitate notation as we develop the right-hand side of (A.3), we write C where the Jacobian is taken at $\frac{x+\xi}{2}$, and substitute $\varepsilon = \frac{x-\xi}{2}$:

$$\left(2C + \left(\frac{\xi - x}{2}\right)^2\right)^{-1} = [(2I + C^{-1}\varepsilon^2)C]^{-1} = \frac{C^{-1}}{2}\left(\frac{1}{I + C^{-1}\varepsilon^2}\right) \tag{A.4}$$

$$= \frac{C^{-1}}{2}\left[I - C^{-1}\frac{\varepsilon^2}{2} + \left(C^{-1}\frac{\varepsilon^2}{2}\right)^2 - \cdots\right] \tag{A.5}$$

$$= \frac{C^{-1}}{2}\left[I - C^{-1}\frac{\varepsilon^2}{2} + O\left(\frac{\varepsilon^4}{4}\right)\right] \tag{A.6}$$

$$= \frac{C^{-1}}{2} - (C^T C)^{-1}\frac{\varepsilon^2}{2} + O\left(\frac{\varepsilon^4}{4}\right) \tag{A.7}$$

$$= \frac{C^{-1}}{2} + O(\|\eta - y\|^2). \tag{A.8}$$

Equality (A.5) is derived with the geometric series summation formula and equality (A.8) is derived by using the bi-Lipschitz-continuity of f. Combining (A.8) with (A.3) completes the proof. □

Appendix B. Proof for Theorem 3.2

Proof. We introduce the change of variables $\hat{y} = y - \bar{y}^{(i)}$ and denote $J_i \equiv J(\bar{y}^{(i)})$, so that (14) becomes

$$W_{ij} = \int_Y \exp\left\{-\frac{\|J_i^{-1}\hat{y}\|^2 + \|J_j^{-1}(\hat{y} - (\bar{y}^{(j)} - \bar{y}^{(i)}))\|^2}{\varepsilon}\right\} d\bar{y}, \tag{B.1}$$

where the ω-weights have canceled. Eq. (B.1) is a convolution of two Gaussians. We apply the Fourier Transform on (B.1) via the convolution theorem [5] and use the Fourier transform of a Gaussian [6]:

$$\mathcal{F}_x[e^{-ax^2}](k) = \sqrt{\frac{\pi}{a}} e^{\frac{-\pi^2 k^2}{a}}, \tag{B.2}$$

to obtain

$$\mathcal{F}_{\bar{y}}[W_{ij}](\xi) = \frac{\pi}{\sqrt{\det(J_i J_i^T)\det(J_j J_j^T)}} \exp\{-\xi^T(J_i J_i^T + J_j J_j^T)\xi\pi^2\varepsilon\}. \tag{B.3}$$

Applying the inverse Transform we obtain

$$W_{ij} = \sqrt{\frac{\pi \det(J_i J_i^T + J_j J_j^T)}{\det(J_i J_i^T)\det(J_j J_j^T)}} \exp\left\{\frac{(\bar{y}^{(j)} - \bar{y}^{(i)})^T[J_i J_i^T + J_j J_j^T]^{-1}(\bar{y}^{(j)} - \bar{y}^{(i)})}{\varepsilon}\right\}. \tag{B.4}$$

Preposition B.1. *The square root factor before the exponent in* (B.4) *can be approximated to a second order by* $\frac{\pi}{\sqrt{\det(J(\bar{y})J(\bar{y})^T)}}$.

Proof. We consider the Taylor expansion of the matrix $C(x) = J(x)J(x)^T$ near $\frac{x+y}{2}$

$$C(x) = C\left(\frac{x+y}{2}\right) + \nabla C\left(\frac{x+y}{2}\right)\left(\frac{x-y}{2}\right) + O((x-y)^2). \tag{B.5}$$

Adding (B.5) to the Taylor expansion of $C(y)$ near $\frac{x+y}{2}$ yields

$$C(x) + C(y) = 2C\left(\frac{x+y}{2}\right) + O((x-y)^2), \tag{B.6}$$

since first order terms cancel. The product $C(x)C(y)$ is approximated by

$$C(x)C(y) = C^2\left(\frac{x+y}{2}\right) - \nabla C^2\left(\frac{x+y}{2}\right)\left(\frac{x-y}{2}\right)^2 + O((x-y)^4). \tag{B.7}$$

We thus obtain that

$$\sqrt{\frac{\det(J_i J_i^T + J_j J_j^T)}{\det(J_i J_i^T)\det(J_j J_j^T)}} \approx \sqrt{\frac{\det(C(\frac{x+y}{2}) + O((x-y)^2))}{\det(C^2(\frac{x+y}{2}) + O((x-y)^2))}} \approx \frac{1}{\sqrt{\det(C(\frac{x+y}{2}))}} \tag{B.8}$$

for x and y sufficiently close. □

Combining Preposition B.1 with (B.4) gives the result (15). □

References

[1] A. Hyvärinen, J. Karhunen, E. Oja, Independent Components Analysis, John Wiley and Sons, New York, 2001.

[2] L.K. Saul, K.Q. Weinberger, J.H. Ham, F. Sha, D.D. Lee, Spectral Methods for Dimensionality Reduction Semisupervised Learning, MIT Press, Cambridge, MA, 2006.

[3] A. Singer, Spectral independent component analysis, Applied and Computational Harmonic Analysis 21 (1) (2006) 128–134.

[4] A. Singer, R.R. Coifman, Non linear independent component analysis with diffusion maps, Applied and Computational Harmonic Analysis 25 (2) (2008) 226–239.

[5] Y. Katznelson, An Introduction to Harmonic Analysis, Dover, New York, 1976.

[6] M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables, Dover, New York, 1972.

[7] N. Aronszajn, Theory of reproducing kernels, Trans. Amer. Math. Soc. 68 (3) (1950) 337–404.

[8] C. Fowlkes, S. Belongie, F. Chung, J. Malik, Spectral grouping using the Nyström method, TPAMI 26 (2) (2004) 214–225.

[9] R.R. Coifman, S. Lafon, Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions, Applied and Computational Harmonic Analysis 21 (1) (2006) 31–52.

[10] S. Lafon, Y. Keller, R.R. Coifman, Data fusion and multicue data matching by diffusion maps, TPAMI 28 (11) (2006) 1784–1797.

[11] A.D. Szlam, M. Maggioni, R.R. Coifman, Regularization on graphs with function-adapted diffusion processes, JMLR 9 (2008) 1711–1739.

[12] L. Rosasco, M. Belkin, E. De Vito, On learning with integral operators, JMLR 11 (2010) 905–934.

[13] Q. Li, D. Omeragic, L. Chou, L. Yang, K. Duong, J. Smits, J. Yang, T. Lau, C. Liu, R. Dworak, V. Dreuillault, H. Ye, New directional electromagnetic tool for proactive geosteering and accurate formation evaluation while drilling, in: SPWLA 46th Annual Logging Symposium, 2005.

[14] R.R. Coifman, S. Lafon, Diffusion maps, Applied and Computational Harmonic Analysis 21 (1) (2006) 5–30.

[15] I.M. Johnstone, On the distribution of the largest eigenvalue in principal components analysis, Annals of Statistics 29 (2001) 295–327.

[16] B. Nadler, Finite Sample Approximation Results for principal component analysis: A matrix perturbation approach, Annals of Statistics 36 (6) (2008) 2791–2817.