We introduce a scheme for the parametrization of an observed measurements set - $\bar{Y}$ into its independent components that enables an efficient extension to a larger set Y. Our work is related to previous work on spectral independent components analysis (ICA), where the non-linear ICA problem is solved via the construction of an anisotropic diffusion kernel whose eigenfunctions comprise the independent components. In our work a different diffusion construct is used, utilizing only the small observed set $\bar{Y}$, that approximates the isotropic diffusion on the parametric manifold $\mathcal{M}_X$ of the full set Y. We employ an extension of the independent components on $\bar{Y}$ to the orthogonal independent components on Y. A validation scheme for our algorithm parameters is also provided, suggesting also a method for the synthesis of new observables by merely using the empirical values of $\bar{X}$. We demonstrate our method on synthetic examples and on semi-supervised classification of electro-magnetic measurements of geological formations.

Anisotropic Diffusion on Sub-Manifolds with Application to Earth Structure Classification

D. Kushnir[†], A. Haddad[†], R. R. Coifman[†]
Technical Report YALEU/DCS/TR-1437
October 20, 2010

[†] Dept. of Mathematics, Yale University, New Haven CT 06511

Approved for public release: distribution is unlimited.
Keywords: *independent components analysis, anisotropic diffusion maps, embedding, graph Laplacian, extension*

1 Introduction

Independent components analysis (ICA) is a fundamental tool in data analysis [1]. Given some high dimensional observed 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. The ICA, as a re-parametrization method, is a useful tool for learning tasks (e.g. classification), since the relations between data points can be computed in a meaningful space.

In order to solve the 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 non-linear transformation $f : X \to Y$. f is an arbitrary non-linear transformation, however, we assume that it is smooth and bi-Lipchitz: there exist a constant $K \geq 0$ such that

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

for all $x^{(i)}, x^{(j)} \in X$. In other words, f is Lipchitz and so is its inverse, thus, both are differentiable 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 by f of the parametric manifold, by using its differential J. 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 approximating the isotropic diffusion on the parametric data manifold $\mathcal{M}_X$, from which the normalized Laplacian of the data graph is constructed. The normalized graph Laplacian asymptotically converges to the Fokker-Planck (FP) operator on the parametric manifold that is separable. Its first non-trivial eigenfunctions are monotonic functions of the independent variables, as guaranteed by Sturm-Liouville 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 observables often arises. A variety of extension techniques via eigenfunctions of data driven matrices have been employed and investigated e.g. [8], [9], [10], [11], [12]. We emphasize the work of [9] which also addressees the issue of the scale of the extension. 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 non-linear ICA [4] in the sense that it computes the ICA of a-priory given data, but also provides an efficient extension for the ICA to any newly observed data. 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]. As a result 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 approximating the eigenfunctions of the FP operator on Y. These eigenfunctions are independent components of Y.

The mapping from the observables space into the independent components space can also 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 non-linear manifold in the observable space may yield erroneous results. We also exploit interpolation in IC-space to compute f for new points on the parametric manifold by merely interpolating known function values. This procedure can be used to synthesize new observables that may also serve to validate the inverse map.

We demonstrate our method on synthetic examples 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 invariancy with respect to different sensors measuring the same physical phenomena. In other words, we observe the same mapping even when the measurements are done by different sensors. The invariancy 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 non-linear 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 observables 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 associat6ed with layered geological models.

2 Problem Definition and Setting

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 non-linear bi-Lipchitz map $f : X \to Y$, such that Y is embedded in a high dimension $Y \subset \mathbb{R}^D$. The intrinsic dimension - d of the data parametric manifold is lower or equal to the dimension of the observable space ($d \leq D$). The variables

x^i are independent $Itô$ processes, given by

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

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

Local covariance and the Jacobian: The $D \times D$ covariance matrix of an observed process is

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

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 $\bar{Y}$ 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 set sub-sampled from the manifold.

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 non-linear map. We take a different strategy then 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 $g : Y \longrightarrow X$ to be the

inverse map of $f : X \longrightarrow Y$, g can be approximated by a Taylor series at the point $\frac{\eta+y}{2}$:

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

where $g^i_j = \frac{\partial g^i}{\partial x^j}$. Therefore the squared Euclidean distance in parametric space can be approximated by

$$\begin{aligned}\left\|\xi - \frac{x+\xi}{2}\right\|^2 = \left\|\frac{\xi-x}{2}\right\|^2 &= \frac{1}{4}\sum_i (\xi^i - x^i)^2 \\ = \quad & \frac{1}{4}\sum_{ijk} g^i_j\left(\frac{\eta+y}{2}\right) g^i_k\left(\frac{\eta+y}{2}\right)(\eta^j - y^j)(\eta^k - y^k) \\ & + \frac{1}{8}\sum_{ijkl} g^i_j\left(\frac{\eta+y}{2}\right) g^i_{kl}\left(\frac{\eta+y}{2}\right)(\eta^j - y^j)(\eta^k - y^k)(\eta^l - y^l) \\ & + O(\|\eta-y\|^4). \end{aligned} \tag{6}$$

A similar approximation at η yields,

$$\begin{aligned}\left\|x - \frac{x+\xi}{2}\right\|^2 = \left\|\frac{x-\xi}{2}\right\|^2 = \left\|\frac{\xi-x}{2}\right\|^2 &= \frac{1}{4}\sum_i (\xi^i - x^i)^2 \\ = \quad & \frac{1}{4}\sum_{ijk} g^i_j\left(\frac{\eta+y}{2}\right) g^i_k\left(\frac{\eta+y}{2}\right)(y^j - \eta^j)(y^k - \eta^k) \\ & - \frac{1}{8}\sum_{ijkl} g^i_j\left(\frac{\eta+y}{2}\right) g^i_{kl}\left(\frac{\eta+y}{2}\right)(\eta^j - y^j)(\eta^k - y^k)(\eta^l - y^l) \\ & + O(\|\eta-y\|^4). \end{aligned} \tag{7}$$

Adding equations (6) and (7), yields a second-order approximation to the Euclidean distance:

$$\|\xi - x\|^2 = (\eta-y)^T[(JJ^T)^{-1}\left(\frac{\eta+y}{2}\right)](\eta-y) + O(\|\eta-y\|^4). \tag{8}$$

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-Lipchitz function $f : X \to Y$, and let $x, y \in X$ such that $y = f(x)$, $\eta = f(\xi)$. Then*

$$(JJ^T)^{-1}\left(\frac{\eta+y}{2}\right) = 2[JJ^T(\eta) + JJ^T(y)]^{-1} + O(\|\eta-y\|^2). \tag{9}$$

Proof is given in Appendix A.

Using Lemma 3.1 in the Eq. (8) yields the second order approximation to the Euclidean distance on X

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

3.2 The diffusion operator on the sub-manifold and its limit operator

Construction of the integral operator. Consider an m-sample of *reference points* $\bar{Y} = \bar{y}^{(1)}, ..., \bar{y}^{(m)}$ from the observed data $y^{(1)}, ..., y^{(N)} \in Y \subseteq \mathbb{R}^D$. $\bar{Y}$ and Y are generated from $\bar{X}$ and X by the same non-linear transformation f as described in section 2. We compute the $m \times N$ affinity matrix between the sample and the set of size N points

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

We compute the matrix

$$W = \omega^{-\frac{1}{2}}(y) A^T A \omega^{-\frac{1}{2}}(y). \tag{12}$$

that comprises the anisotropic diffusion weights between on the points in the reference set $\bar{Y}$ via the points in Y. The density normalization used in (12) is applied by using the vector of sums of A's column entries: $\omega(y) = diag(A^T \mathbf{1})$, where $\mathbf{1}$ is a column vector with all entries equal to 1. This particular normalization corresponds to the approximation of the FP operator and its eigenfunctions [15], that is pursued further below.

Remark 3.2. *The transformation $\omega^{-\frac{1}{2}}(y)A$ can be viewed as a map from $\bar{Y}$ to Y with the density measure $\omega(y)$.*

The following result approximates the matrix (12) 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 3.3. *The kernel*

$$W_{ij} = \int_Y \exp\left\{-\frac{\|J^{-1}(\bar{y}^{(i)})(\bar{y}^{(i)} - y)\|^2 + \|J^{-1}(\bar{y}^{(j)})(\bar{y}^{(j)} - y)\|^2}{\varepsilon}\right\} \omega^{-\frac{1}{2}}(y)\omega^{-\frac{1}{2}}(y)\omega(y)dy \tag{13}$$

corresponding to the matrix (14) can be approximated to a second order by

$$W_{ij} = \frac{\pi}{det(J^T(\tilde{y})J(\tilde{y}))} \exp\left\{-\frac{\|[J(y^{(i)}) + J(y^{(j)})]^{-1}(y^{(j)} - y^{(i)})\|^2}{\varepsilon}\right\} \tag{14}$$

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 Lemma 3.3, it is straightforward to show that as $\varepsilon \to 0$ and $m \to \infty$ the discrete operator (12) converges to the integral operator

$$\frac{1}{m}\sum_{i=1}^{m} W_{ij} q_{\bar{Y}}(\bar{y}^{(j)}) \to \int_{\bar{Y}} \exp\left\{\frac{\|[J(y)+J(\bar{y}^{(i)})]^{-1}(y-\bar{y}^{(i)})\|^2}{\varepsilon}\right\} p_{\bar{Y}}(y) q_{\bar{Y}}(y) dy \qquad (15)$$

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 $\frac{1}{det(J(\tilde{y})J(\tilde{y})^T)}$ approximates the density $p_{\bar{Y}}$ near $\tilde{y}$. The 2nd order approximation suggested in Eq. (10) can be used now so that (15) is equal to the integral on the parametric space $\bar{X}$

$$\int_{\bar{X}} \exp\left\{-\frac{\|\bar{x}^{(i)}-x\|^2 + \|\bar{x}^{(i)}-x\|^4}{2\varepsilon}\right\} p_{\bar{X}}(x) q_{\bar{X}}(x) dx, \qquad (16)$$

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

$$L = D^{-1}W - I \qquad (17)$$

constructed from W (15), with $D = \{d_{ii}\}_{i=1}^{m}$, converges to the backward FP operator $\mathcal{L}$ on $\bar{X}$ [4], [15]:

$$\mathcal{L}q = \triangle q - \nabla U \cdot \nabla q, \qquad (18)$$

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:

$$\mathcal{L} = \triangle + 2\nabla \log p_{\bar{X}} \nabla = \sum_i \left(\frac{\partial^2}{\partial^2 \bar{x}^i} + 2\frac{\partial log p^i(\bar{x}^i)}{\partial \bar{x}^i}\frac{\partial}{\partial \bar{x}^i}\right) = \sum_i \mathcal{L}_i, \qquad (19)$$

where $\mathcal{L}_i$ corresponds to the 1-dimensional backward FP operator on the data. The eigenfunctions $\Phi_d = [\varphi_1, ..., \varphi_d]$ of the operator (19) are monotonic functions of the processes x^i as guarantied by Sturm-Liouville theory. Thus one can use them to *re-parameterize* 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 non-linear transformation f, up to a local scaling.

4 Extension

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 (11). We start with discussing the SVD in the setting of reproducing kernel Hilbert space [7]. The SVD provides the algebraic relation between the matrix A (eq. 11) and W (eq. 12): 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 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 straight forward 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 $m-sample$ $\bar{Y}=\{y^{(i)}\}_{i=1}^{m}$ of $Y=\{y^{(i)}\}_{i=1}^{N}$ such that $\bar{Y}\subset Y$, and let μ be a measure on Y. Let $\hat{W}$ be a symmetric, semi-positive definite, bounded kernel $\hat{W}: Y\times Y\rightarrow\mathbb{R}$. 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 refer to A defined in (11) as if it was normalized: $A=\omega^{-\frac{1}{2}}A$. Then the density-normalized operator $A: L^2(\bar{Y},\omega^{-\frac{1}{2}}d\mu)\rightarrow\mathcal{H}$ (known as the extension operator) and its adjoint $A^*:\mathcal{H}\rightarrow L^2(\bar{Y},\omega^{-\frac{1}{2}}d\mu)$ (restriction operator) can be used to construct the Hilbert-Schmidt operator $A^*A: L^2(\bar{Y},\omega d\mu)\rightarrow L^2(\bar{Y},\omega d\mu)$, defined in (12). The operator A^*A is compact, self adjoint and has a discrete sequence of eigenvalues and eigenvectors.

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

$$
\begin{aligned}
A\varphi_j &= \sqrt{\lambda}\psi_j \quad j=1,...,k\\
A\varphi_j &= 0 \quad j=k+1,...,m\\
A^*\psi_j &= \sqrt{\lambda}\psi_j \quad j=1,...,k\\
A^*\psi_j &= 0 \quad j=k+1,...,m,
\end{aligned}
$$

and

$$
\begin{aligned}
A^*A\varphi_j &= \lambda\varphi_j \quad j=1,...,k\\
A^*A\varphi_j &= 0 \quad j=k+1,...,m\\
AA^*\psi_j &= \lambda\psi_j \quad j=1,...,k\\
AA^*\psi_j &= 0 \quad j=k+1,...,N.
\end{aligned}
$$

The following observations are motivated by the SVD formulation:

1. A^*A and AA^* are positive, self-adjoint matrices.

2. The spectra of of A^*A and AA^* are the same, and

$$\psi_j(y) = \frac{1}{\sqrt{\lambda_j}} \int_{\bar{Y}} A(y,s)\varphi_j(s)d\mu(s), \quad y \in Y. \tag{20}$$

3. The eigenfunctions of A^*A and AA^* coincide up to a scaling factor on $\bar{Y}$.

4. ψ are the extension of φ outside the set $\bar{Y}$.

5. The extension preserves the orthonormality of the eigenfunctions.

We conclude that the extension Ψ_d of the independent components Φ_d on $\bar{Y}$ yields independent components on Y because each $\psi_i,\ i = 1, ..., d$ is a weighted combination of the independent ϕ_i. The monotonicity property is preserved in Ψ since the weights are based on a 2nd-order approximation of the type (8) to the Euclidean distances on the parametric manifold. Finally, we emphasis the computational gain obtained by computing Ψ via (20) instead of using a direct method for the eigen-decomposition of the matrix AA^*.

4.2 Synthetic example

We demonstrate the spectral ICA and its eigenvector extension with a synthetic example. The example employs the non-linear 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

$$\begin{aligned} r_c^{(i)} &= r^{(i)} + \sqrt{dt}w^{(i)} \\ \theta_c^{(i)} &= \theta^{(i)} - \tfrac{\theta^{(i)}}{2}dt + \sqrt{dt}w^{(i)}, \end{aligned} \tag{21}$$

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

$$\bar{y}^{(i)} = (r^{(i)}cos(\theta^{(i)}), r^{(i)}sin(\theta^{(i)})) \tag{22}$$

and

$$y_c^{(i)} = (r_c^{(i)}cos(\theta_c^{(i)}), r_c^{(i)}sin(\theta_c^{(i)})) \tag{23}$$

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. 1-A,B, respectively.

Next, we construct the weight matrix W (Eq. (12)), using the local sample covariance matrix to approximate JJ^T via Eq. (5), and $\varepsilon = 0.2$. The right

eigenvectors $\{\phi\}_{i=1}^{2}$ 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 $\phi_0 = [1, 1, ..., 1]$ and thus disregarded. The next two eigenvectors are monotonic functions of r and θ [3]. We construct the embedding of $\bar{Y}$ using $\Phi_2 : \bar{y}^{(i)} \rightarrow [\varphi_1(\bar{y}_1^{(i)}), \varphi_2(\bar{y}_2^{(i)})]$. 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. 1 C, D the embedding of the reference points is illustrated, where the color coding corresponds to the true parameters r and θ. The color coding suggest 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.

4.3 Validation via synthesis of new observations

Extension of functions on the data can be done in various ways (see for example [9], [10], [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

$$\Psi_d^{-1}(x) = \sum_{k:\bar{x}^{(k)} \in N_x} c_k \bar{y}^{(k)}, \tag{24}$$

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 points $\tilde{x}$ in the re-parametrization Ψ_d and synthesize a corresponding observable by using the interpolation formula in (24) 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.

Equation (24) 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 $\Psi_d = [\psi_1(y), ..., \psi_d(y)]$, 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

$$e = \|y - \Psi_d^{-1}(x)\|. \tag{25}$$

We demonstrate the optimization of the parameter ε for the synthetic example

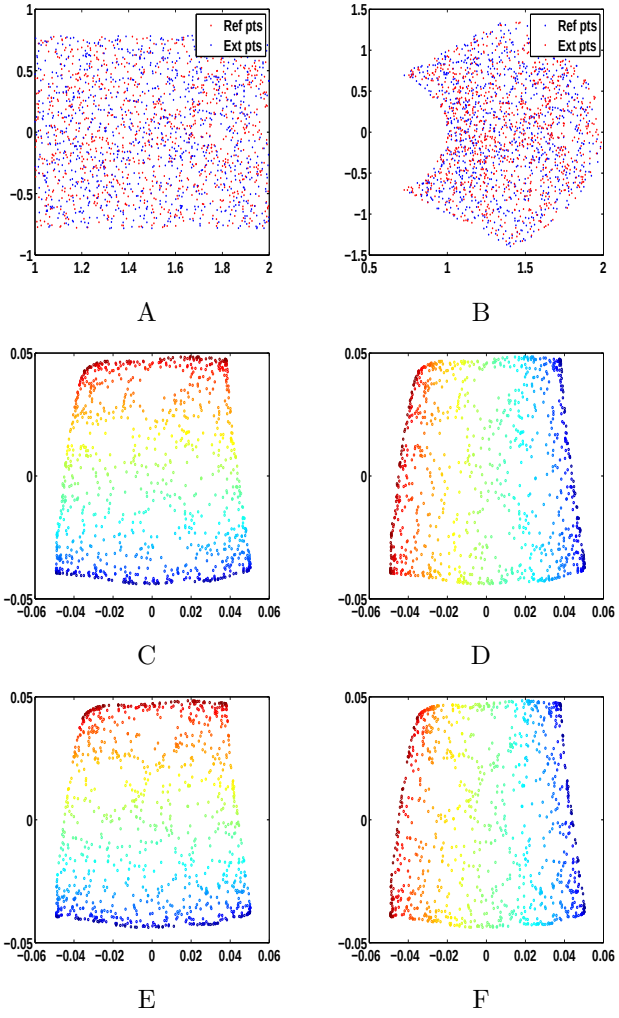

Figure 1: A Synthetic example. **A,B:** 2D illustration of data in parametric space (A), and in observed space (B). **C,D:** embedding of reference points into Φ_2 with color coding corresponding to r (A), and θ (B). **E,F:** Embedding of the extended points into Ψ_2 with color coding corresponding to r (E), and θ (F).

in section 4.2. The interpolation coefficients c_k in (24) are computed by

$$c_k(x) = \frac{exp\left(-\frac{\|x-\bar{x}^{(k)}\|^2}{\sigma_{\Psi_d(y)}}\right)}{\sum_{i:\bar{x}^{(i)}\in N_{\bar{X}}} exp\left(-\frac{\|x-\bar{x}^{(i)}\|^2}{\sigma_{\Psi_d(y)}}\right)}. \tag{26}$$

The mean error (25) is computed for 100 randomly sampled points $y^{(i)}$ in the synthetic example described in section 4.2. $\sigma_{\Psi_d(y^{(i)})}$ 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 (25) with respect to ε only. The minimal error $e = 0.05$ is attained at $\varepsilon = 0.2$. The error is of order $O(\varepsilon^2)$ as implied from the approximation (10).

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

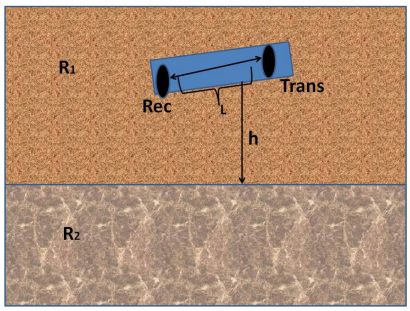

Figure 2: Two-layers parametric model. Layers resistivity is denoted by R, whereas the distance between the tool layer and the boundary is denoted by h.

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 measuremnts and the corresponding geological parameters. One approach that could be used is based on the 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 non-linearity, the Euclidean distances do not reflect any informative proximity between the data points, unless the manifold is oversampled. 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 d of the tool to the other layer and the layer resistivity R_2. We use two tool channels: one with spacing of 96" and 100kHz, and the other with spacing 34"

and 400kHz. 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

$$\bar{y}^{(i)} = (Att_1, PS_1, Att_2, PS_2)^T \tag{27}$$

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

$$\bar{x}^{(i)} = (R_2, h, R_1)^T \tag{28}$$

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, for each $\bar{y}^{(i)}$ corresponds 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 equation (11)). We next construct the 961×961 matrix W (Eq. (12)). 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

$$L = D^{-1}W - I. \tag{29}$$

The first eigen-pair is disregarded, whereas the next two eigenvectors approximate the eigenvectors of the FP operator and thus are monotonic functions of d and R_2, i.e. the independent components. We construct the re-parametrization of $\bar{Y}$ using $\Phi_2 : \bar{y}^{(i)} \to [\varphi_1(\bar{y}_1^{(i)}), \varphi_2(\bar{y}_2^{(i)})]$: $\bar{y}^{(i)}$ is mapped to a vector in $\mathbb{R}^2$ containing the $i-th$ coordinate of the first two (non-trivial) eigenvectors of (29). Next, we construct the extension to Y, $\Psi_2 : y^{(i)} \to [\psi_1(y_1^{(i)}), \psi_2(y_2^{(i)})]$, by using the extension formula (20).

We plot the reference points $\bar{X}$ in the 2D parametric space in figure 3-A, and the complete data set parameters - X in Fig. 3-B. The x-axis corresponds to R_2 and the y-axis corresponds to h. To demonstrate the non-linearity of f, the observed reference points are plotted in Fig. 3-C,D. In Fig. 3-E and F we plot the reference points re-parametrization Φ_2 and its extension to Ψ_2, respectively.

Extension error. Given the true parameters of 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

$$\tilde{x}_{new}^{(i)} = \sum_{k:\Psi_d(\bar{y}^{(k)}) \in N_i} c_i \bar{x}^{(k)}, \tag{30}$$

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

$$err_{prm}(\Psi_d(y^{(i)})) = \|x^{(i)} - \tilde{x}_{new}^{(i)}\|^2. \tag{31}$$

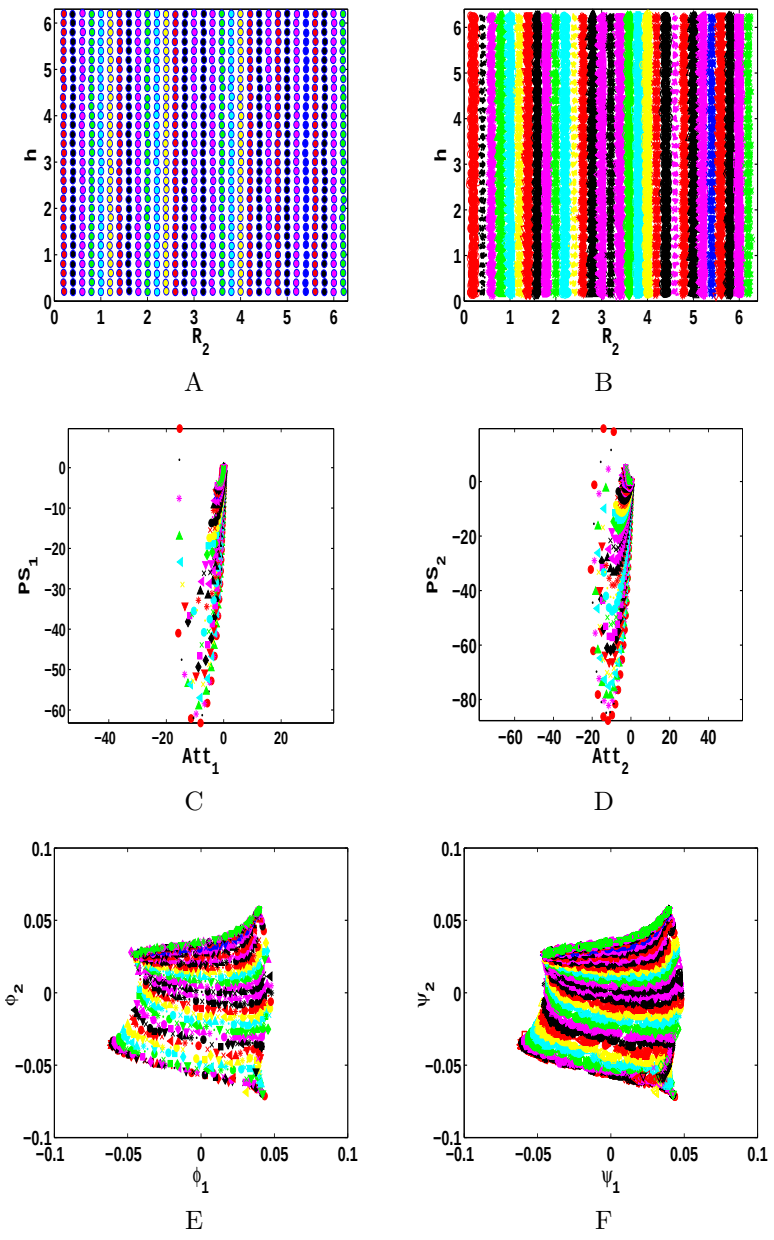

Figure 3: Embedding and Extension of EM Data. **A,B:** The 2D parametric space of d and R_2 sampled. $\bar{X}$ (A), X (B). **C,D:** Observed data $\bar{Y}$. 2D illustration of observed data in 96" channel (C). 34" channel (D). **E,F:** Data re-parametrization. Embedding of reference points $\bar{Y}$ by Φ_2 (E). The extension Ψ_2 (F).

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

$$\hat{x}^{(i)}_{new} = \sum_{k:\bar{y}^{(k)} \in N_i} c_i \bar{x}^{(k)}, \tag{32}$$

where c_i are linear interpolation coefficients computed in Y. The corresponding error is measured by

$$err_{obs}(y^{(i)}) = \|x^{(i)} - \hat{x}^{(i)}_{new}\|^2. \tag{33}$$

In the following we show that we can reduce the number of reference points and still obtain a low re-parametrization error - $err(y^{(i)})$ (31) then the interpolation error in the observable space (33). For the purpose of this demonstration $\bar{Y}$ is sub-sampled uniformly for fewer and fewer points. The remaining set $Y \setminus \bar{Y}$ is used as a pool of points for testing our extension. In Fig. 5.1 the error is plotted against the rate of sampling of the two parameters R_2 and h. In figure 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. On the other hand, the classification error using Euclidean metric in Y is almost uniformly distributed on the interval $[0, 2]$.

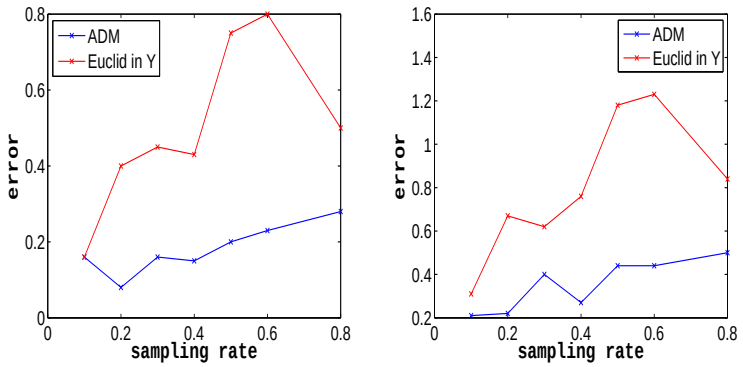

Figure 4: Extension errors of ICA extension and a Euclidean-based lookup table. Left: mean errors of resistivity - R_2. Right: mean errors of distance to boundary - h.

Feature space selection. We demonstrate that our ICA method can be used as a tool for the selection of the features 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 100kHz and 400kHz. The second data set is generated using one channel of 96 inches with 100kHz and another channel of 96 inches with 400kHz. We also generate a data set

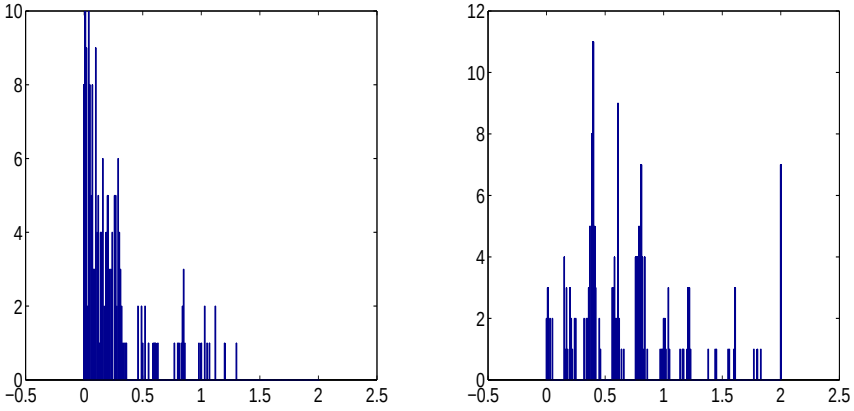

Figure 5: Histograms of extension error of ICA extension and a Euclidean-based lookup table. Left: errors of h for our method. Right: error of h for the Euclidean metric in Y.

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 non-linear 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 (32) - 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 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.

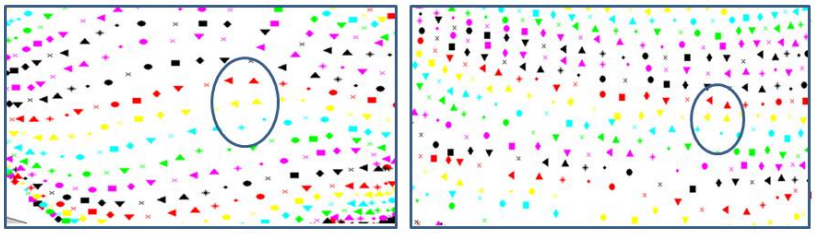

Figure 6: Similar maps of different channel sets into their independent components. Left: embedding of channels of distance 96 inches with frequencies 100kHz and 400kHz. Right: embedding of channels of distance 96 inches and 34 inches with frequencies 100kHz and 400kHz, respectively. A sampled region of correspondence is circled in both embedding.

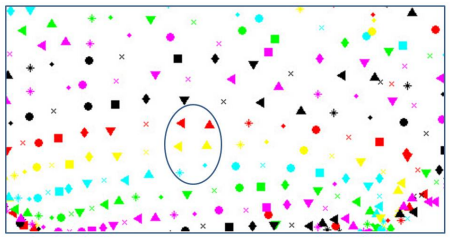

Figure 7: Similar maps of different channel sets into their independent components: embedding of all channels. A sampled region of correspondence is circled in both embedding.

6 Conclusion

We have described a method for the computation and extension of independent components of high dimensional data generated by a non-linear 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 observables. 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 observeable space.

7 Acknowledgment

The authors would like to thank Nick Bennett and Schlumberger for their support in this work.

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

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

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 $\varepsilon = \frac{x-\xi}{2}$:

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

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

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

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

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

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

B Proof for Lemma 3.3

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

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

where the ω-weights have canceled. Equation (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]:

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

to obtain

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

Applying the inverse transform we obtain

$$W_{ij} = \sqrt{\frac{\pi det(J_i J_i^T + J_j J_j^T)}{det(J_i J_i^T) det(J_j J_j^T)}} \exp\left\{ \frac{\|[J_i J_i^T + J_j J_j^T]^{-1}(y^{(j)} - y^{(i)})\|^2}{\varepsilon} \right\}. \tag{B.4}$$

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

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

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

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

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

for x and y sufficiently close. □

Combining preposition B.1 with (B.4) gives the result (14). □

References

[1] A. Hyvärinen, J. Karhunen, E. Oja, Indepndent 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] Abramowitz, M. and Stegun, I. A., Handbook of Mathematical Functrions 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, and H. Ye, New Directional Electromagnetic Tool For Proactive Geosteering And Accurate Formation Evaluation While Drilling, SPWLA 46th Annual Logging Symposium, 2005

[14] D. Omeragic, T. Habashy, C. Esmersoy, Q. Li, J. Seydoux, J. Smits, J. R. Tabanou, Real-Time Interpretation Of Formation Structure From Directional EM Measurements, SPWLA 47th Annual Logging Symposium, 2006

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