ALTERNATING DIFFUSION FOR COMMON MANIFOLD LEARNING WITH APPLICATION TO SLEEP STAGE ASSESSMENT

Roy R. Lederman[1]*, Ronen Talmon*[2]*,Hau-tieng Wu*[3]*, Yu-Lun Lo*[4]*, Ronald R. Coifman*[1]

[1]Department of Mathematics, Yale University, New Haven CT, USA.
[2]Department of Electrical Engineering, Technion - Israel Institute of Technology, Haifa, Israel.
[3]Department of Mathematics, University of Toronto, Toronto, Ontario, Canada.
[4]Department of Thoracic Medicine, Chang Gung Memorial Hospital
and School of Medicine, Chang Gung University, Taipei, Taiwan.

ABSTRACT

In this paper, we address the problem of multimodal signal processing and present a manifold learning method to extract the common source of variability from multiple measurements. This method is based on alternating-diffusion and is particularly adapted to time series. We show that the common source of variability is extracted from multiple sensors as if it were the only source of variability, extracted by a standard manifold learning method from a single sensor, without the influence of the sensor-specific variables. In addition, we present application to sleep stage assessment. We demonstrate that, indeed, through alternating-diffusion, the sleep information hidden inside multimodal respiratory signals can be better captured compared to single-modal methods.

Index Terms— Common variable, alternating-diffusion, diffusion maps, multimodal, sleep

1. INTRODUCTION

In recent years, there has been a growing effort to develop analysis methods based on low dimensional geometry driven by measurements. Among the emerging methodologies is *Manifold Learning*, e.g., ISOMAP [1], locally linear embedding (LLE) [2], Hessian and Laplacian Eigenmaps [3, 4], and Diffusion Maps [5, 6]. The core of manifold learning often resides in the construction of a kernel based on local connections between the samples of a signal. Then, these local connections are aggregated into a global representation, usually through an eigenvector problem. Under the appropriate conditions, manifold learning yields a low dimensional representation of the main structures of the data. However, since the measurements are typically influenced by various nuisance and interfering factors, manifold learning might capture a combination of the desired as well as the undesired structures.

A particular problem of interest is multimodal signal processing, where the same physical phenomenon is measured using multiple types of instruments or sensors. As a result, each set of related measurements of the same phenomenon has a different geometric structure, depending on the specific instrument. This poses a significant obstacle to existing methods, since each additional sensor introduces additional redundant sources of variability, while the ultimate goal is to describe the physical phenomenon independently of the way it is measured. Here, we address the multimodal problem from a (data-driven) manifold learning standpoint. The multimodal case presents a challenge to manifold learning, since multiple sensors often increase the number of undesired structures. Nevertheless, multiple sensors enable us to recover a more reliable description of the measured physical phenomenon if we can extract the common source of variability from the different measurements while suppressing the sensor-specific factors.

The widely-used Canonical Correlation Analysis (CCA) is designed to extract the common source of variability from multiple measurements [7] . However, the applicability of this approach is limited when the relations between the common source of variability and the measurements are nonlinear. Several nonlinear methods to analyze data from multiple sensors have been proposed based on kernel CCA [8] or [9], combining affinities [10], constructing Markov chains [11, 12], aggregating features [13], and cross-diffusion [14]. However, these methods are not designed to distinguish between the common variable and the sensor-related variables.

Recently, we have presented a manifold learning method based on alternating-diffusion to extract the common source of variability from multimodal measurements [15]. In this paper, we extend this method to time series, in which we assume that the driving sources of variability are independent. We show that the common source of variability is extracted by this method from multiple sensors as if it were the only source of variability, extracted by a standard manifold learning method from a single sensor, without the influence of the sensor-specific variables. In addition, we present application to sleep stage assessment. We demonstrate that, indeed, through alternating-diffusion, the sleep information hidden inside multichannel signals can be better captured compared to single-channel methods.

2. PROBLEM FORMULATION

Consider three hidden random variables $(X, Y, Z) \sim \pi(X, Y, Z)$, from the (possibly high dimensional) spaces $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$. Following standard practice (e.g., principal component analysis (PCA) [16]), we assume that the variables are statistically independent, i.e., their joint probability density $\pi(X, Y, Z)$ is given by

$$\pi(X, Y, Z) = \pi_x(X)\pi_y(Y)\pi_z(Z), \tag{1}$$

where π_x, π_y, and π_z are the marginal densities of X, Y, and Z, respectively, and they have unit variances.

These hidden variables are measured through two random variables $S = g(X, Y)$ and $R = h(X, Z)$ from the (possibly high

Ronen Talmon acknowledges the support by the European Union's - Seventh Framework Programme (FP7) under Marie Curie Grant No. 630657. Yu-Lun Lo acknowledges the support by Taiwan National Science Council grant 101-2220-E-182A-001 and 102-2220-E-182A-001.

dimensional) spaces $\mathcal{S}$ and $\mathcal{R}$, where g and h are two unknown measurement functions with suitable regularity conditions.

From n realizations of the system, we have n triplets of the hidden variables $\{(x_i, y_i, z_i)\}_{i=1}^n$ and n pairs of corresponding measurements $\{(s_i, r_i)\}_{i=1}^n$. We refer to s_i and r_i as the measurement in Sensor 1 and the measurement in Sensor 2, respectively. We note that both s_i and r_i are functions of x_i, to which we refer as the common variable, whereas y_i and z_i are the sensor-specific variables.

Our goal is to recover a parameterization of the samples of the common variable $\{x_i\}_{i=1}^n$ from the pairs of measurements $\{(s_i, r_i)\}_{i=1}^n$.

3. MAHALANOBIS DISTANCE

Under the assumption that the hidden variables X, Y and Z are independent, Singer and Coifman [17] showed that the Euclidean distance between the variables can be locally estimated from the sensor measurements to the second order by:

$$\|(x_i, y_i) - (x_j, y_j)\|_2^2 = \|s_i - s_j\|_{\text{md}}^2 + \mathcal{O}\left(\|s_i - s_j\|_2^4\right) \quad (2)$$

where $\|\cdot\|_{\text{md}}$ is the *Mahalanobis distance* given by

$$\|s_i - s_j\|_{\text{md}}^2 = \frac{1}{2}(s_i - s_j)^T(\mathbf{C}_i^\dagger + \mathbf{C}_j^\dagger)(s_i - s_j) \quad (3)$$

and $\mathbf{C}_i^\dagger$ is the pseudo-inverse of the covariance matrix $\mathbf{C}_i$ of S evaluated locally at i. In case of time series, the covariance $\mathbf{C}_i$ can be estimated from samples in a short-time window around i by the sample covariance [18, 19]. We remark that the approximation (2) is obtained by the Taylor expansion of the function g and will be combined with an exponentially decaying kernel, where the approximation error becomes negligible. For more details, we refer the readers to [17, 20]. Similarly, we have

$$\|(x_i, z_i) - (x_j, z_j)\|_2^2 = \|r_i - r_j\|_{\text{md}}^2 + \mathcal{O}\left(\|r_i - r_j\|_2^4\right) \quad (4)$$

with an analogous definition of the Mahalanobis distance based on samples from Sensor 2. For simplicity, we omit the sensor index from the Mahalanobis distance notation.

4. ALTERNATING DIFFUSION

We begin by separately constructing affinity matrices $\mathbf{W}^{(\mathbf{s})}$ and $\mathbf{W}^{(\mathbf{r})}$ from the measurements of each sensor based on a Gaussian function and the Mahalanobis distance as follows:

$$\begin{aligned} W_{ij}^{(\mathbf{s})} &= \exp\left(-\|s_i - s_j\|_{\text{md}}^2/\varepsilon\right) \\ W_{ij}^{(\mathbf{r})} &= \exp\left(-\|r_i - r_j\|_{\text{md}}^2/\varepsilon\right) \end{aligned} \quad (5)$$

for all $i, j = 1, \ldots, n$, where ε is the kernel's scale. For simplicity, we use a single scale, set to be the square of the median of the distances. However, we can use different scales for the different measurements.

Then, two diffusion operators (Markov matrices) $\mathbf{K}^{(\mathbf{s})}$ and $\mathbf{K}^{(\mathbf{r})}$ are created by dividing each column of the kernel by its sum as

$$K_{ij}^{(\mathbf{s})} = W_{ij}^{(\mathbf{s})} / \sum_{l=1}^n W_{lj}^{(\mathbf{s})}; K_{ij}^{(\mathbf{r})} = W_{ij}^{(\mathbf{r})} / \sum_{l=1}^n W_{lj}^{(\mathbf{r})} \quad (6)$$

Based on $\mathbf{K}^{(\mathbf{s})}$ and $\mathbf{K}^{(\mathbf{r})}$, the alternating propagation from the i-th sample is defined as a sequence of vectors $\{v_t\}_{t=0}^\infty$ such that

$$v_{t+1} = \begin{cases} \mathbf{K}^{(\mathbf{s})} v_t, & t = 2m \\ \mathbf{K}^{(\mathbf{r})} v_t, & t = 2m+1 \end{cases} \quad (7)$$

for every integer $m \geq 0$, where the initial vector v_0 is a vector of dimensionality n of all zeros, except for the i-th position, which is assigned the value 1, i.e., $v_0 = (0, \ldots, 0, 1, 0, \ldots, 0)^T$, where we omit the dependency on i for simplicity. It follows that the even steps of the propagation in (7) can be restated as

$$v_{2m} = \mathbf{K}^m v_0. \quad (8)$$

where $\mathbf{K}$ is the alternating-diffusion (Markov) matrix

$$\mathbf{K} = \mathbf{K}^{(\mathbf{r})}\mathbf{K}^{(\mathbf{s})}. \quad (9)$$

We define the diffusion distance between Sample i and Sample j based on the alternating-diffusion as the following Euclidean distance

$$d_t(i, j) = \|v_t - u_t\|_2, \quad (10)$$

where v_t and u_t are vectors in the propagation sequences (defined in (7)) from Sample i and Sample j, respectively. The diffusion distance has been shown to be a powerful metric of comparing samples that is invariant to small topological distortions and moderate noise [6, 21]. While the Euclidean distance compares two individual samples, the diffusion distance integrates other samples and measures the relationship of the two samples via the entire sample set. We will show in Section 5 that the new alternating-diffusion operator $\mathbf{K}$ can be viewed as an "effective" diffusion operator, and that it captures the structure of the common variable and ignores the variables specific to either sensor. It follows that the diffusion distances computed from the accessible measurements (s_i, r_i) are equivalent to diffusion distances computed from data, where the common variable is the only source of variability.

The diffusion distance is typically approximated through low-dimensional features, traditionally, using the eigenvalue decomposition (EVD) or the singular value decomposition (SVD) of $\mathbf{K}$. Here, the SVD is applied to $\mathbf{K}^t$, and the right-singular vectors $\{\psi_\ell\}_{\ell=1}^n$ associated with the large singular values $\{\lambda_\ell\}_{\ell=1}^n$ are used to define an embedding. Let $\Psi_{i,t}$ denote the d-dimensional embedding of the pair (s_i, r_i), defined by

$$\Psi_{i,t} : (s_i, r_i) \mapsto (\lambda_1\psi_1(i), \lambda_2\psi_2(i), \ldots, \lambda_d\psi_d(i))^T. \quad (11)$$

We remark that a possible refinement of the algorithm is a second application of diffusion maps with a kernel based on the diffusion distances [15], i.e., the Euclidean distances between the low-dimensional features (11).

The extension of the algorithm to measurements from multiple sensors is straightforward. Let $\mathbf{K}^{(1)}, \ldots, \mathbf{K}^{(m)}$ denote the kernels corresponding to measurement sets from m sensors, computed similarly to (5) and (6). Then, the alternating-diffusion Markov matrix is given by

$$\mathbf{K} = \mathbf{K}^{(m)} \cdots \mathbf{K}^{(1)}. \quad (12)$$

5. ANALYSIS

Following standard practice in diffusion geometry, the analysis presented here is in the continuous domain [5, 6]. We define the continuous counterpart of the kernel of Sensor 1 by

$$K^{(\mathbf{s})}((x, y), (x', y')) = \frac{1}{\omega^{(\mathbf{s})}(x', y')} e^{-\|s - s'\|_{\text{md}}^2/\epsilon}, \quad (13)$$

where $s = g(x, y)$, $s' = g(x', y')$, and $\omega^{(\mathbf{s})}(x', y')$ is a normalization factor defined (using the independence of X and Y) by

$$\omega^{(\mathbf{s})}(x', y') = \int_{\mathcal{X},\mathcal{Y}} e^{-\|s'' - s'\|_{\text{md}}^2/\epsilon} \pi_x(x'')\pi_y(y'')dx''dy'' \quad (14)$$

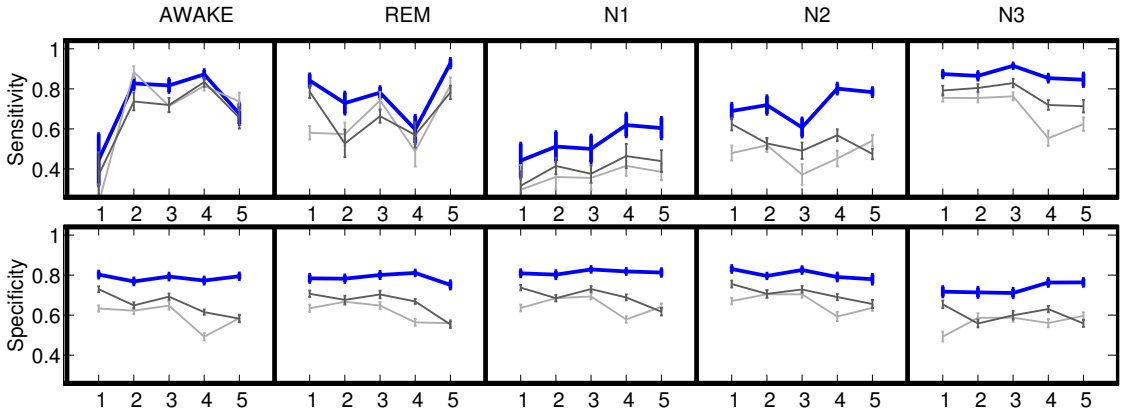

Fig. 1. The sensitivity and the specificity of the sleep stage classification per individual. The thin gray curves represent the classification results based on the single channels (light gray - airflow, dark gray - abdominal motion), and the thick blue curve represents the classification results from the alternating diffusion.

where $s'' = g(x'', y'')$.

By combining (2) with the exponentially decaying Gaussian, the kernel is asymptotically separable with respect to the independent variables, i.e.,

$$K^{(\mathbf{s})}((x,y),(x',y')) = K^{(\mathbf{x})}(x,x')K^{(\mathbf{y})}(y,y') \tag{15}$$

where

$$K^{(\mathbf{x})}(x,x') = \frac{1}{\omega^{(\mathbf{x})}(x')} e^{-\|x-x'\|^2/\epsilon} \tag{16}$$

$$K^{(\mathbf{y})}(y,y') = \frac{1}{\omega^{(\mathbf{y})}(y')} e^{-\|y-y'\|^2/\epsilon} \tag{17}$$

and the normalization factors are

$$\omega^{(\mathbf{x})}(x') = \int_{\mathcal{X}} e^{-\|x''-x'\|^2/\epsilon} \pi_x(x'')dx'' \tag{18}$$

$$\omega^{(\mathbf{y})}(y') = \int_{\mathcal{Y}} e^{-\|y''-y'\|^2/\epsilon} \pi_y(y'')dy'' \tag{19}$$

For a function $p : \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \to \mathbb{R}$, the continuous diffusion operation $D^{(\mathbf{s})}$ of Sensor 1 is given by

$$\left(D^{(\mathbf{s})}p\right)(x,y,z) = \int_{\mathcal{X},\mathcal{Y},\mathcal{Z}} K^{(\mathbf{s})}((x,y),(x',y'))p(x',y',z') \times \pi(x',y',z')dx'dy'dz' \tag{20}$$

The continuous diffusion operator $D^{(\mathbf{r})}$ of Sensor 2 is defined in an analogous manner.

By combining the single-modal diffusion operators $D^{(\mathbf{s})}$ and $D^{(\mathbf{r})}$, we construct an alternating-diffusion operator. For a function $p_0 : \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \to \mathbb{R}$, we define a sequence of *propagated functions* $\{p_t\}_{t=0}^{\infty}$ by

$$p_{t+1}(x,y,z) = \begin{cases} \left(D^{(\mathbf{s})}p_t\right)(x,y,z), & \forall t = 2m \\ \left(D^{(\mathbf{r})}p_t\right)(x,y,z), & \forall t = 2m+1 \end{cases} \tag{21}$$

where $m \geq 0$ is a nonnegative integer and $p_0(x,y,z)$ is typically a function with support only at an initial point (x,y,z), which increases along the diffusion process.

Thus far, we have presented the continuous counterpart of the discrete diffusion described in Section 4, which can be computed directly from the measurements. Now, we introduce the *effective function* $p_t^{(\mathbf{e})}(x) : X \to \mathbb{R}$, which is defined by

$$p_t^{(\mathbf{e})}(x) = \int_{\mathcal{Y},\mathcal{Z}} p_t(x,y,z)\pi_y(y)\pi_z(z)dydz. \tag{22}$$

Theorem 1. *For every $t \geq 1$, the effective function $p_{t+1}^{(\mathbf{e})}(x)$ is related to the preceding effective function $p_t^{(\mathbf{e})}(x)$ by*

$$p_{t+1}^{(\mathbf{e})}(x) = \left(D^{(\mathbf{e})}p_t^{(\mathbf{e})}\right)(x) \tag{23}$$

where $D^{(\mathbf{e})}$ is the effective diffusion operator defined by

$$\left(D^{(\mathbf{e})}p^{(\mathbf{e})}\right)(x) = \int_{\mathcal{X}} K^{(\mathbf{x})}(x,x')\, p^{(\mathbf{e})}(x')\pi_x(x')dx' \tag{24}$$

Proof. The proof is obtained by substituting (24), (22), (21), (20), and (15) into (23) and changing the order of integration. □

In other words, Theorem 1 implies that the alternating-diffusion of propagating functions on the measurements is effectively a diffusion on "marginal-like" functions defined solely on the hidden common variable.

The effective functions $p_t^{(\mathbf{e})}$ and the effective operator $D^{(\mathbf{e})}$ are merely formal objects that are computed directly. Nevertheless, we will relate the effective functions to the accessible propagating function through a distance metric.

Theorem 2. *Suppose that $\{p_t(x,y,z)\}$ and $\{q_t(x,y,z)\}$ are two sequences of propagated functions as defined in (21), and that $\{p_t^{(\mathbf{e})}(x)\}$ and $\{q_t^{(\mathbf{e})}(x)\}$ are the corresponding sequences of effective propagated functions as defined in (22). Then, for $t > 0$*

$$\|p_t^{(\mathbf{e})}(x) - q_t^{(\mathbf{e})}(x)\|_M = \|p_{t+1}(x,y,z) - q_{t+1}(x,y,z)\|_\pi, \tag{25}$$

where $\|\cdot\|_\pi$ is defined by

$$\|p_t(x,y,z)\|_\pi = \left(\int_{\mathcal{X},\mathcal{Y},\mathcal{Z}} p_t^2(x,y,z)\pi(x,y,z)dxdydz\right)^{1/2} \tag{26}$$

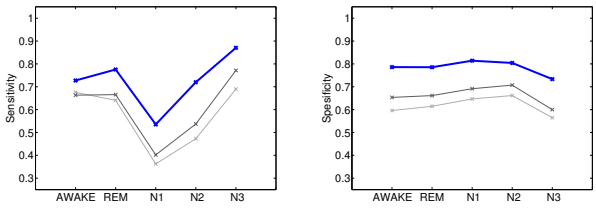

Fig. 2. The average sensitivity and specificity of the sleep stage classification over the five individuals. The thin gray curves represent the classification results based on the single channels (light gray - airflow, dark gray - abdominal motion), and the thick blue curve represents the classification results from the alternating diffusion.

and $\|\cdot\|_M$ *is defined by*

$$\|p_t^{(\mathbf{e})}(x)\|_\pi = \Big(\int_{\mathcal{X},\mathcal{X}} p_t^{(\mathbf{e})}(x) M(x,x') p_t^{(\mathbf{e})}(x') \times \pi_x(x)\pi_x(x')dxdx'\Big)^{1/2} \quad (27)$$

$$M(x,x') = \int_{\mathcal{X}} K^{(\mathbf{x})}\left(x'',x\right) K^{(\mathbf{x})}\left(x'',x'\right) \pi_x(x'')dx'' \quad (28)$$

Proof. We refer the readers to the proof of Theorem 5 in [15], where the result is obtained by plugging-in (15). □

In other words, the effective operator $D^{(\mathbf{e})}$ generates a diffusion on $\mathcal{X}$, and although the sequences of effective functions are not accessible directly, the diffusion distance associated with $D^{(\mathbf{e})}$ can be computed from the data.

6. EXPERIMENTAL RESULTS

We applied the alternating diffusion to sleep recordings to assess the sleep stage of an individual. Physiologically, sleep is divided into two broad stages: rapid eye movement (REM), and non-rapid eye movement (NREM) [22]. The NREM stage is further divided into shallow sleep (stages N1 and N2) and deep sleep (stage N3). To assess sleep, various signals are typically recorded. Among these signals, EEG signals are the most concentrated, since the clinically acceptable stage of the sleep is majorly determined by reading the recorded EEG based on the R&K criteria [23,24]. However, it is well known that sleep is a global and systematic behavior not localized solely in the brain. In particular, there have been several studies of the sleep stage analysis based on the respiratory signal [25–29].

In [30], we considered the model in which there exists an evolutionary hidden process restricted to a low-dimensional Riemannian manifold and governing the respiratory signals. The idea that lies behind the model is that the accessible respiratory signal is a measurement of the neural system controlling the sleep cycle. While it can be affected by numerous factors related to the measurement modality (e.g., measurements of airflow or chest movements), the equipment (e.g., the type of sensors and their exact positions) and noise, the true intrinsic variable we have interest in is the intrinsic state controlling the respiratory signal.

In this paper, we apply the alternating diffusion to extract the common source of variability in abdominal motions, which are recorded by the piezo-electric bands, as well as in airflow, which is measured using thermistors and nasal pressure, both at the sampling rate 100 Hz. These two measurements were chosen, since they were shown to be the least informative with regard to sleep among the respiratory and EEG signals [30]. The primary idea is that the common variable of the different respiratory signals (measured by different types of instruments) recovers true physiological hidden process that is well correlated with the sleep stage, and simultaneously, gets rid of nuisance factors, which are sensor-specific.

Five subjects without sleep apnea were chosen for this study. The demographic characteristics of these individuals fall within the norm. We used 6 hour length recordings, which were performed in the sleep center at Chang Gung Memorial Hospital (CGMH), Linkou, Taoyuan, Taiwan. The institutional review board of the CGMH approved the study protocol (No. 101-4968A3) and the enrolled subjects provided written informed consent. See [30] for more details regarding the experimental setting and the collected data.

We consider five sleep stages in this study: Awake; REM; N1; N2; N3. We computed three sets of data-driven embeddings from the two channels of respiratory signals: airflow and abdominal motion. Each of the single-channel recording was preprocessed by applying the Scattering Transform [31], which was shown to improve the regularity and stability of the signal with respect to various deformations [31]. Then, we computed the kernel for each channel according to (5) and (6) and separately applied the SVD to each of the kernels to obtain its diffusion maps embedding [6]. Finally, we applied the alternating diffusion method by combining the two channels; the two kernels were combined to a single alternating diffusion kernel according to (9) and applied SVD to obtain the embedding (11) representing the common variable of the channels.

In order to evaluate the quality of the single and multiple modal embeddings, the embedded samples are used as input to multiclass support vector machine (SVM) with the radial basis function (RBF) to classify sleep stages. To prevent over-fitting and confirm the classification result, we performed repeated random sub-sampling validation 25 times and evaluated the average. We randomly partitioned the data into training dataset and validation dataset – the training dataset comprised 80% of the samples and the rest are used to form the validation dataset. The trained classifier based on the training dataset is used to predict the sleep stages of the validation dataset.

Fig. 1 depicts the sensitivity and the specificity of the sleep stage classification per individual. The thin gray curves represent the classification results based on the single channels and the thick blue curve represents the classification results from the alternating diffusion. We observe that for most individuals, the classification results based on combined information is superior. To further demonstrate the effectiveness of the alternating-diffusion in extracting a true physiological variable, Fig. 2 depicts the average sensitive and specificity over the five individuals. The results imply that extracting the common source of variability indeed "filters" out nuisance factors and enhances the correspondence to the true physiological variable (the sleep stage), which is hidden in the respiratory signals.

The main purpose of this application is to demonstrate the ability of the alternating-diffusion to extract the common variable from real-world multimodal measurements; we showed that the application enhances the correspondence to the true physiological variable. Since the alternating-diffusion method is completely data-driven, the results imply the existence of information of the sleep stage underlying the respiratory signals. Moreover, the improvement with respect to single modal embeddings suggests that each sensor measure other factors besides the sleep. Although the results do not improve the state of the art, they show that sleep stage identification is possible from the respiratory signal, and the results are comparable with the results obtained based on EEG signals. We remark that these results need to be validated on a larger dataset with more individuals.

7. REFERENCES

[1] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 260, pp. 2319–2323, 2000.

[2] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 260, pp. 2323–2326, 2000.

[3] D. L. Donoho and C. Grimes, "Hessian eigenmaps: New locally linear embedding techniques for high-dimensional data," *Proc. Nat. Acad. Sci.*, vol. 100, pp. 5591–5596, 2003.

[4] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, pp. 1373–1396, 2003.

[5] R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, "Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps," *Proc. Nat. Acad. Sci.*, vol. 102, no. 21, pp. 7426–7431, May 2005.

[6] R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 21, pp. 5–30, Jul. 2006.

[7] Harold Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321, Dec. 1936.

[8] C. Lai, P. L. Fyfe, "Kernel and nonlinear canonical correlation analysis," *International Journal of Neural Systems*, vol. 10, no. 5, pp. 365–377, 2000.

[9] B. Boots and G. J. Gordon, "Two-manifold problems with applications to nonlinear system identification," *In Proc. 29th Intl. Conf. on Machine Learning (ICML)*, 2012.

[10] V. R. de Sa, P. W. Gallagher, J. M. Lewis, and V. L. Malave, "Multi-view kernel construction," *Machine Learning*, vol. 79, no. 1-2, pp. 47–71, May 2010.

[11] D. Zhou and C. J. C. Burges, "Spectral clustering and transductive learning with multiple views," *Proceedings of the 24th international conference on Machine learning (ICML)*, pp. 1159–1166, 2007.

[12] H.-C. Huang, Y.-Y. Chuang, and C.-S. Chen, "Affinity aggregation for spectral clustering," *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 773–780, June 2012.

[13] Y. Keller, R. R. Coifman, S. Lafon, and S. W. Zucker, "Audio-visual group recognition using diffusion maps," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 403–413, Jan. 2010.

[14] B. Wang, J. Jiang, W. Wang, Z.-H. Zhou, and Z. Tu, "Unsupervised metric fusion by cross diffusion," *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 2997–3004, June 2012.

[15] R. R. Lederman and R. Talmon, "Common manifold learning using alternating-diffusion," *submitted, Tech. Report YALEU/DCS/TR1497*, 2014.

[16] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 3, pp. 611–622, 1999.

[17] A. Singer and R. Coifman, "Non-linear independent component analysis with diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 25, pp. 226–239, 2008.

[18] R. Talmon and R. R. Coifman, "Empirical intrinsic geometry for nonlinear modeling and time series filtering," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 110, no. 31, pp. 12535–12540, 2013.

[19] R. Talmon and R. R. Coifman, "Intrinsic modeling of stochastic dynamical systems using empirical geometry," *to appear in Applied and Computational Harmonic Analysis (ACHA), Tech. Report YALEU/DCS/TR1467*, 2014.

[20] R. Talmon, S. Mallat, Z. Hitten, and R. R. Coifman, "Manifold learning for latent variable inference in dynamical systems," *submitted, Tech. Report YALEU/DCS/TR1491*, 2014.

[21] N. El Karoui and H.-T. Wu, "Connection graph Laplacian methods can be made robust to noise," *to appear in Annals of Statistics, arXiv:1405.6231*, 2014.

[22] T. Lee-Chiong, *Sleep Medicine: Essentials and Review*, Oxford, 2008.

[23] A. Rechtschaffen and A. Kales, *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*, Washington: Public Health Service, US Government Printing Office, 1968.

[24] C. Iber, S. Ancoli-Isreal, A.L. Chesson Jr., and S.F. Quan, *The AASM Manual for Scoring of Sleep and Associated Events-Rules: Terminology and Technical Specification*, American Academy of Sleep Medicine, 2007.

[25] G. S. Chung, B. H. Choi, K. K. Kim, Y. G. Lim, J. W. Choi, D.-U. Jeong, and K.-S. Park, "Rem sleep classification with respiration rates," in *Information Technology Applications in Biomedicine, 2007. ITAB 2007. 6th International Special Topic Conference on*, 2007, pp. 194–197.

[26] G. Guerrero-Mora, P. Elvia, A.M. Bianchi, J. Kortelainen, M. Tenhunen, S.L. Himanen, M.O. Mendez, E. Arce-Santana, and O. Gutierrez-Navarro, "Sleep-wake detection based on respiratory signal acquired through a pressure bed sensor," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 2012, pp. 3452–3455.

[27] J. Sloboda and M. Das, "A simple sleep stage identification technique for incorporation in inexpensive electronic sleep screening devices," in *Aerospace and Electronics Conference (NAECON), Proceedings of the 2011 IEEE National*, 2011, pp. 21–24.

[28] H.-T. Wu, "Instantaneous frequency and wave shape functions (I)," *Appl. Comput. Harmon. Anal.*, vol. 35, pp. 181–199, 2013.

[29] Y.-C. Chen, M.-Y. Cheng, and H.-T. Wu, "Nonparametric and adaptive modeling of dynamic seasonality and trend with heteroscedastic and dependent errors," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 76, pp. 651–682, 2014.

[30] H.-T. Wu, R. Talmon, and Y.-L. Lo, "Assess Sleep Stage by Modern Signal Processing Techniques," *to appear in IEEE Trans. Biomed. Eng., arXiv:1410.1013*, 2015.

[31] S. Mallat, "Group invariant scattering," *Pure and Applied Mathematics*, vol. 10, no. 65, pp. 1331–1398, 2012.