



## Alternating diffusion maps for multimodal data fusion

Ori Katz<sup>\*,a</sup>, Ronen Talmon<sup>a</sup>, Yu-Lun Lo<sup>b</sup>, Hau-Tieng Wu<sup>c,d,e</sup>

<sup>a</sup> Viterbi Faculty of Electrical Engineering, Technion, Israel Institute of Technology, Israel

<sup>b</sup> Department of thoracic medicine, Chang Gung Memorial Hospital, Chang Gung University, School of Medicine, Taipei, Taiwan

<sup>c</sup> Department of Mathematics, University of Toronto, Ontario, Canada

<sup>d</sup> Department of Mathematics and department of statistical science, Duke university, Durham, NC, USA

<sup>e</sup> Mathematics Division, National Center for Theoretical Sciences, Taipei, Taiwan



### ARTICLE INFO

#### Keywords:

Common manifold learning  
Multimodal sensor fusion  
Nonlinear-filtering  
Diffusion maps  
Alternating diffusion maps

### ABSTRACT

The problem of information fusion from multiple data-sets acquired by multimodal sensors has drawn significant research attention over the years. In this paper, we focus on a particular problem setting consisting of a physical phenomenon or a system of interest observed by multiple sensors. We assume that all sensors measure some aspects of the system of interest with additional sensor-specific and irrelevant components. Our goal is to recover the variables relevant to the observed system and to filter out the nuisance effects of the sensor-specific variables. We propose an approach based on manifold learning, which is particularly suitable for problems with multiple modalities, since it aims to capture the intrinsic structure of the data and relies on minimal prior model knowledge. Specifically, we propose a nonlinear filtering scheme, which extracts the hidden sources of variability captured by two or more sensors, that are independent of the sensor-specific components. In addition to presenting a theoretical analysis, we demonstrate our technique on real measured data for the purpose of sleep stage assessment based on multiple, multimodal sensor measurements. We show that without prior knowledge on the different modalities and on the measured system, our method gives rise to a data-driven representation that is well correlated with the underlying sleep process and is robust to noise and sensor-specific effects.

### 1. Introduction

Often, when measuring a phenomenon of interest that arises from a complex dynamical system, a single data acquisition method is not capable of capturing its entire complexity and characteristics, and it is usually prone to noise and interferences. Recently, due to technological advances, the use of multiple types of measurement instruments and sensors have become more and more popular; nowadays, such equipment is smaller, less expensive, and can be mounted on every-day products and devices more easily. In contrast to a single sensor, multimodal sensors may capture complementary aspects and features of the measured phenomenon, and may enable us to extract a more reliable and detailed description of the measured phenomenon.

The vast progress in the acquisition of multimodal data calls for the development of analysis and processing tools, which appropriately combine data from the different sensors and handle well the inherent challenges that arise. One particular challenge is related to the heterogeneity of the data acquired in the different modalities; datasets acquired from different sensors may comprise different sources of variability, where only few are relevant to the phenomenon of interest.

This particular challenge as well as many others have been the subject of many studies. For a recent comprehensive reviews, see [1–3].

In this paper we consider a setting in which a physical phenomenon is measured by multiple sensors. While all sensors measure the same phenomenon, each sensor consists of different sources of variability; some are related to the phenomenon of interest, possibly capturing its various aspects, whereas other sources of variability are sensor-specific and irrelevant. We present an approach based on manifold learning, which is a class of nonlinear data-driven methods, e.g. [4–7], and specifically, we use the framework of diffusion maps (DM) [8]. On the one hand, manifold learning is particularly suitable for problems with multiple modalities since it aims to capture the intrinsic geometric structure of the underlying data and relies on minimal prior model knowledge. This enables to handle multimodal data in a systematic manner, without the need to specially tailor a solution for each modality. On the other hand, applying manifold learning to data acquired in multiple (multimodal) sensors may capture undesired/nuisance geometric structures as well. Recently, several manifold learning techniques for multimodal data have been proposed [9–12]. In [9], the authors suggest to concatenate the samples acquired by different sensors

\* Corresponding author.

E-mail address: [orikats@tx.technion.ac.il](mailto:orikats@tx.technion.ac.il) (O. Katz).

into unified vectors. However this approach is sensitive to the scaling of each dataset, which might be especially diverse among datasets acquired by different modalities. To alleviate this problem, it is proposed in [10] to use DM to obtain “standardized” representation of each dataset separately, and then to concatenate these “standardized” representations into the unified vectors. Despite handling better multimodal data, this concatenation scheme does not utilize the mutual relations and co-dependencies that might exist between the datasets.

While methods such as those presented in [9,10,12] take into account all the measured information, the methods presented in [11,13–15] use local kernels to implement nonlinear filtering. Specifically, following a recent line of study in which multiple kernels are constructed and combined [16–19], in [13,14], it was shown that a method based on alternating applications of diffusion operators extracts only the common source of variability among the sensors, while filtering out the sensor-specific components. Therefore we choose to establish our method based on DM which relies on those theoretical foundations. For other nonlinear methods, such as local CCA [11] and kernel CCA [15], more efforts are needed to better understand their theoretical foundation, yet, they may also be used as an alternative to DM and alternating diffusion (AD) and empirically tested as well. The shortcoming of alternating applications of diffusion operators arises when having a large number of sensors; often, sensors that measure the same system capture different information and aspects of that system. As a result, the common source of variability among all the sensors captures only a partial or empty look of the system, and important relevant information may be undesirably filtered out.

Here, we address the tradeoff between these two approaches. That is, we aim to maintain the relevant information captured by multiple sensors, while filtering out the nuisance components. Since the relevance of the various components is unknown, our main assumption is that the sources of variability which are measured only in a single sensor, i.e., sensor-specific, are nuisance. Conversely, we assume that components measured in two or more sensors are of interest. Importantly, such an approach implements implicitly a smart “sensor selection”; “bad” sensors that are, for example malfunctioned and measure only nuisance information, are automatically filtered out. These assumptions stem from the fact that the phenomenon of interest is global and not specific to one sensor. We propose a nonlinear filtering scheme, in which only the sensor-specific sources of variability are filtered out while the sources of variability captured by two or more sensors are preserved.

Based on prior theoretical results [13,14], we show that our scheme indeed accomplishes this task. We illustrate the main features of our method on a toy problem. In addition, we demonstrate its performance on real measured data in an application for sleep stage assessment based on multiple, multimodal sensor measurements. Sleep is a global phenomenon with systematic physiological dynamics that represents a recurring non-stationary state of mind and body. Sleep evolves in time and embodies interactions between different subsystems, not solely limited in the brain. Thus, in addition to the well-known patterns in electroencephalogram (EEG) signals, its complicated dynamics are manifested in other sensors such as sensors measuring breathing patterns, muscle tones and muscular activity, eyeball movements, etc. Each one of the sensors is characterized by different structures and affected by numerous nuisance processes as well. In other words, while we could extract the sleep dynamics by analyzing different sensors, each sensor captures only part of the entire sleep process, whereas it introduces modality artifacts, noise, and interferences. We show that our scheme allows for an accurate systematic sleep stage identification based on multiple EEG recordings as well as multimodal respiration measurements. In addition, we demonstrate its capability to perform sensor selection by artificially adding noise sensors.

The remainder of the paper is organized as follows. In Section 2 we present a formulation for the common source extraction problem and present an illustrative toy problem. In Section 3, a brief review for the

method proposed in [13,14] is outlined, and then, a detailed description and interpretation of the proposed scheme are presented. In Section 4, we first demonstrate the capabilities of the proposed scheme on the toy problem introduced in Section 2. Then, in Section 5, we demonstrate the performance in sleep stage identification based on multimodal measured data recorded in a sleep clinic. Finally, in Section 6, we outline several conclusions.

## 2. Problem setting

Consider a system driven by a set of  $K$  hidden random variables  $\Theta = \{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(K)}\}$ , where  $\theta^{(k)} \in \mathbb{R}^{d_k}$ . The system is measured by  $M$  observable variables  $s^{(m)}$ ,  $m = 1, \dots, M$ , where each sensor has access to only a partial view of the entire system and its driving variables  $\Theta$ . To formulate it, we define a “sensitivity table” given by the binary matrix  $\mathbf{S} \in \mathbb{Z}_2^{K \times M}$ , indicating the variables sensed by each observable variable. Specifically, the  $(k,m)$ th element in  $\mathbf{S}$  indicates whether the hidden variable  $\theta^{(k)}$  is measured by the observable variable  $s^{(m)}$ . It should be noted that this binary notation is a simplification that does not take into account “soft degrees of observability”, e.g., the nonlinearity of the observation function, measurement noise, etc. Theoretically, when we have sufficient data, the algorithm is guaranteed to work for any bilipschitz observation function. Yet, when the amount of data is limited, these degrees of observability are dominated by the Lipschitz constants of the observation function and by the signal-to-noise ratio. Some of these observability aspects were addressed in [20], and further quantification is left for future work. The observable variables are therefore given by

$$s^{(m)} = h_m(\Theta^{(m)}, \mathbf{n}^{(m)}) \in \mathbb{R}^{D_m} \quad (1)$$

where  $h_m(\cdot)$  is a bilipschitz observation function,  $\mathbf{n}^{(m)} \in \mathbb{R}^{p_m}$  are hidden random variables captured only by the  $m$ th observable variable, and  $\Theta^{(m)}$  is the subset of driving hidden variables of interest sensed by  $s^{(m)}$ , given by

$$\Theta^{(m)} = \{\theta^{(k)} \mid \forall k, S_{k,m} = 1\} \subseteq \Theta, m = 1, \dots, M \quad (2)$$

The random hidden variables  $\mathbf{n}^{(m)}$  are *sensor-specific* (associated only with the  $m$ th observer). They are conditionally independent given the hidden variables of interest and will be assumed as noise/nuisance variables. We further assume that each random hidden variable in  $\Theta$  is measured by at least two observable variables, such that  $\sum_{m=1}^M S_{k,m} \geq 2$  for each  $k = 1, \dots, K$ . As a result, we refer to the hidden variables  $\theta^{(k)}$  in  $\Theta$  as *common variables*.

In order to simplify the notation, we denote the subset of all hidden variables (both common and sensor-specific) measured by the  $m$ th observable by  $\mathcal{S}^{(m)} = \{\Theta^{(m)}, \mathbf{n}^{(m)}\}$ . Furthermore, we assume that the dimensions of the observations and the hidden variables satisfy

$$D_m \geq \sum_{\theta^{(k)} \in \Theta^{(m)}} (d_k + p_k), k = 1, 2, \dots, M \quad (3)$$

i.e., the observations are in higher dimension than the hidden common and nuisance variables.

An observation of the system denoted as  $(s_i^{(1)}, s_i^{(2)}, \dots, s_i^{(M)})$  is associated with a realization of the hidden variables  $\Theta_i = (\theta_i^{(1)}, \dots, \theta_i^{(K)})$  and realizations of the  $M$  hidden nuisance variables  $(\mathbf{n}_i^{(1)}, \dots, \mathbf{n}_i^{(M)})$ . Given  $N$  observation samples  $\{(s_i^{(1)}, s_i^{(2)}, \dots, s_i^{(M)})\}_{i=1}^N$ , our goal is to obtain a parametrization for the underlying realizations of the common hidden random variables  $\{(\theta_i^{(1)}, \dots, \theta_i^{(K)})\}_{i=1}^N$  while filtering out the nuisance variables  $\{(\mathbf{n}_i^{(1)}, \dots, \mathbf{n}_i^{(M)})\}_{i=1}^N$ . We note that the observations index  $i$  may represent the time index in case of time series.

### 2.1. Illustrative toy problem

We illustrate the problem setting using the following toy example. Consider six rotating arrows captured in simultaneous snapshots by

Download English Version:

<https://daneshyari.com/en/article/6937920>

Download Persian Version:

<https://daneshyari.com/article/6937920>

[Daneshyari.com](https://daneshyari.com)