Many natural and artificial high-dimensional time series are often controlled by a set of lower-dimensional independent factors. In this paper anisotropic diffusion is combined with local dynamical models to provide intrinsic global modeling that reveals these factors. The obtained model is shown to be invariant to the measuring equipment and can be efficiently extended. These two properties are paramount for sequential processing and provide a foundation for probabilistic analysis. The widely applicable approach is demonstrated on nonlinear tracking problems based on both simulated and recorded data.

Differential Stochastic Sensing: Intrinsic Modeling of Random Time Series with Applications to Nonlinear Tracking

Ronen Talmon and Ronald R. Coifman
Research Report YALEU/DCS/TR-1451
Yale University
February 14, 2012

Approved for public release: distribution is unlimited.
Keywords: *intrinsic modeling, anisotropic diffusion, nonlinear tracking*

1 Introduction

Natural and artificial high-dimensional data is often highly structured and does not fill uniformly the high-dimensional space. In recent years there has been much progress in the development of new methods for parameterizing and embedding high-dimensional data in a low-dimensional space [1] [2] [3] [4] [5] [6]. The nonlinear independent component analysis (NLICA) proposed in ref. [7] is of particular interest since the data is assumed to be inaccessible and can be observed only via unknown nonlinear functions. By integrating local principal component analysis with diffusion maps the NLICA approach provides modeling of the underlying parametric manifold, whereas the classical manifold learning methods provide parametrization of the observable manifold.

In many natural and artificial systems, the unknown functions that map the low-dimensional data into a subset of high-dimensional observations may be stochastic. For example, neurons in the auditory or visual systems map arbitrary external stimuli into random sequences of electrical pulses. In addition, a random measurement noise usually corrupts the observations. In [8], we tackled the stochastic nature of measurements by assuming Gaussian models and estimating the second order statistics of the data. However, in many non-Gaussian cases, the second order statistics does not contain all the information and might not be sufficient.

In this paper, we present a framework for inferring the independent controlling factors of high-dimensional time series. A "differential stochastic sensing" approach is proposed to model noisy measurements based on local density estimates and anisotropic diffusion. This approach is specially adapted to cases where the observation functions are stochastic. We show that the obtained intrinsic modeling is invariant under different measurement schemes and is noise resilient. Moreover, it implicitly encodes the temporal dynamics of the data and can be efficiently extended. Hence, it provides a foundation for sequential processing that is illustrated on nonlinear tracking problems. Observe that the proposed tracking procedure is independent of the model construction.

Our implicit assumption is that the processing of time series occurs in three time scales. The micro time scale is the measurements domain, where we assume that the sampling rate is sufficiently high to enable us to estimate the varying densities in short windows. The mezzo scale is the density domain. In this domain, the noise statistics are assumed to be constant or slowly changing. Finally, the macro scale is the anisotropic diffusion domain model, where all the available information is aggregated into a global dynamical model.

2 Dynamical and Measurement Models

Let $\mathbf{y}_t$ denote an n-dimensional observation process in time index t, drawn from a time varying probability density function (pdf), denoted by $p_{y;t;\theta}(\mathbf{y})$. Consider a model in which the pdf is controlled by a d-dimensional process $\boldsymbol{\theta}_t$ consisting of independent factors. The dynamics of the controlling factors are described by normalized independent Itô processes as follows

$$d\theta_t^i = a^i(\theta_t^i)dt + dw_t^i,\ i = 1, \ldots, d, \tag{1}$$

where a^i are unknown drift coefficients and $\dot{w}_t^i$ are independent white noises. In discrete time, Eq. (1) can be re-written as

$$\theta_t^i = f^i(\theta_{t-1}^i, \dot{w}_t^i),\ i = 1, \ldots, d, \tag{2}$$

where f represents the temporal propagation of the independent factors.

Let $\mathbf{z}_t$ denote a noisy version of the observation $\mathbf{y}_t$, which is given by

$$z_t^j = g^j(\mathbf{y}_t, \mathbf{v}_t),\ j = 1, \ldots, n, \tag{3}$$

where g is an arbitrary (possibly nonlinear) measurement function and $\mathbf{v}_t$ is a corrupting n-dimensional measurement noise drawn from an unknown stationary pdf $p_v(\mathbf{v})$ and independent of $\mathbf{y}_t$.

Our goal is to recover the original controlling factors by constructing an intrinsic model based merely on noisy observations.

2.1 Relationship to Bayesian Filtering

In the formulation of the traditional Bayesian filtering problem, the measured process observes the underlying controlling process, which is often referred to as state, via a deterministic possibly nonlinear function [9]. Thus, in our setting it is equivalent to a special case where $p_{y,t,\theta}(\mathbf{y}_t) \triangleq \delta(\mathbf{y}_t - \boldsymbol{\theta}_t)$, and $\delta(\cdot)$ is the Dirac delta function. If we further restrict f and g to be linear and $\dot{\mathbf{w}}_t$ and $\mathbf{v}_t$ to be independent white Gaussian noises, we obtain the linear-Gaussian case which can be optimally solved by the Kalman Filter [10]. Many extensions of the Kalman Filter to nonlinear models have been developed in the last decades. The most popular among them are the Extended Kalman Filter [9] and the Unscented Kalman Filter [11]. In addition, sequential Monte Carlo integration methods have been proposed, which have the advantage of not being restricted to linearity or Gaussianity (See refs. [12] and [13] and the reference therein). All these methods require at least partial knowledge of the dynamical and the measurement models and the noise statistics, whereas in this work we infer the models blindly based solely on observations.

3 Toy Example

We aim to track a moving object on a 3-dimensional sphere. Since the radius of the sphere is fixed, we assume that the movement of the object is controlled by two independent factors $\boldsymbol{\theta}_t = [\theta_t^1; \theta_t^2]$: the horizontal azimuth angle θ_t^1 and the vertical elevation angle θ_t^2. Suppose the temporal propagation of the angles mimics a motion of a particle in a potential field. Thus, the angles evolve in time according to the following Langevin equation

$$\dot{\theta}_t^i = -\nabla U(\theta_t^i) + \dot{w}_t^i,\ i = 1, 2, \tag{4}$$

where $\dot{w}_t^i$ are independent white Gaussian noises, and U is the potential field. See Fig. 1 for an examples of such 2-dimensional trajectory. Let $\mathbf{x}(\boldsymbol{\theta}_t)$ denote the 3-dimensional coordinates of the object position on the sphere. By assuming that the center of the sphere

is located at the origin of the coordinate system, the position of the object is given by

$$\begin{aligned} x^1(\boldsymbol{\theta_t}) &= r\cos(\theta_t^1)\sin(\theta_t^2) \\ x^2(\boldsymbol{\theta_t}) &= r\sin(\theta_t^1)\sin(\theta_t^2) \\ x^3(\boldsymbol{\theta_t}) &= r\cos(\theta_t^2), \end{aligned}$$

where r is the radius of the sphere. We consider three measurement schemes:

1. The movement of the object is measured in 3 sensors positioned in $\mathbf{x}_j, j = 1,2,3$ outside the sphere (see Fig. 2 for the setup illustration). The sensors detect the object and fire spikes through a spatial point process in a varying rate which depends on the proximity of the object to the sensors. The spikes are fired by each sensor according to a Poisson distribution with rate $\lambda^j(\boldsymbol{\theta_t}) = \exp\{-\|\mathbf{x}_j - \mathbf{x}(\boldsymbol{\theta_t})\|\}$. This enables us to test the challenging non-Gaussian class of tracking models. We obtain three spike trains y_t^j in which the temporal rate is higher when the object is closer to the sensor. The output of each sensor is corrupted by additive noise and is given by
$$z_t^j = g^j(y_t, v_t) = y_t^j + v_t^j,\ j = 1,2,3,$$
where v_t^j is a spike train drawn from a Poisson distribution with a fixed rate λ_v^j.

2. Similarly to the second scheme, each sensor fires spikes randomly according to the proximity of the object. However, in this scheme we simulate sensors with unreliable clocks. We measure the time interval between two consecutive spikes, which is given by $z_t^j = y_t^j + v_t^j$. Suppose y_t^j is drawn from exponential distribution with a rate parameter $\lambda^j(\boldsymbol{\theta}_t) = \exp\{-\|\mathbf{x}_j - \mathbf{x}(\boldsymbol{\theta}_t)\|\}$, and suppose v_t^j is drawn from a fixed normal distribution representing the clock inaccuracy.

3. In this scheme we consider a measurement of a different nature. We use three sensors that measure the location of the source directly, i.e.
$$z_t^j = x_t^j + v_t^j,\ j = 1,2,3,$$
where v_t^j is an additive Gaussian white noise.

4 Intrinsic Modeling

4.1 Histograms

Let $p_{z;t;\theta}(\mathbf{z}_t)$ denote the pdf of the measured process at time index t.

Lemma 1. *The pdf of the measured process is a linear transformation of the pdf of the clean observation component.*

The proof of Lemma 1 is straightforward. By relying on the independence of $\mathbf{y}_t$ and $\mathbf{v}_t$, the pdf of the measured process is given by

$$\begin{aligned} p_{z;t;\theta}(\mathbf{z}_t) &= \int\limits_{g(\mathbf{y}_t,\mathbf{v}_t)=\mathbf{z}_t} p(\mathbf{y}_t,\mathbf{v}_t)d\mathbf{y}_t d\mathbf{v}_t \\ &= \int\limits_{g(\mathbf{y}_t,\mathbf{v}_t)=\mathbf{z}_t} p_{y;t;\theta}(\mathbf{y}_t)p_v(\mathbf{v}_t)d\mathbf{y}_t d\mathbf{v}_t. \end{aligned} \tag{5}$$

In the common additive measurement noise case, i.e., $g(\mathbf{y}_t, \mathbf{v}_t) = \mathbf{y}_t + \mathbf{v}_t$, only a single solution $\mathbf{v}(\mathbf{z}_t) = \mathbf{z}_t - \mathbf{y}_t$ to $g(\mathbf{y}_t, \mathbf{v}_t) = \mathbf{z}_t$ exists. Thus, $p_{z;t;\theta}(\mathbf{z}_t)$ in Eq. (5) becomes a linear convolution.

By Lemma 1 and by the definition of the pdf of the clean observation component $p_{y;t;\theta}(\mathbf{y}_t)$ we get the following result.

Corollary 1. *The pdf of the measured process is a nonlinear function of the controlling factors.*

In practice, we can only estimate the time varying pdf of the measurements by calculating histograms in short-time windows. Since histograms are obtained by integrating the pdf they can be seen as linear transformations. Let $\mathbf{h}_z(\boldsymbol{\theta}_t)$ be the local histogram of the measured process $\mathbf{z}_t$ at time t, whose elements are ideally given by

$$h_z^k(\boldsymbol{\theta}_t) = \frac{1}{(2\delta)^n} \int_{z_c^1-\delta}^{z_c^1+\delta} \cdots \int_{z_c^n-\delta}^{z_c^n+\delta} p_{z;t;\theta}(\mathbf{z}) d\mathbf{z}, \tag{6}$$

where $\mathbf{z}_c$ is a predefined vector consisting of the centers of the high-dimensional histogram bins of width 2δ.

One of the emerging challenges is the computation of high-dimensional histograms, especially when the dimension of the measurement is high. In order to reduce the dimension without corrupting the information the data may preprocess by applying random filters.

4.2 Noise-robust Metric

The local histograms are viewed as feature vectors for each measurement, i.e.,

$$\mathbf{z}_t \mapsto \mathbf{h}_z(\boldsymbol{\theta}_t).$$

Then, for each feature vector, we compute the local covariance matrix in a time interval of length L according to

$$\boldsymbol{\Sigma}_t = \frac{1}{L} \sum_{s=t}^{t-L+1} (\mathbf{h}_z(\boldsymbol{\theta}_s) - \boldsymbol{\mu}_t)(\mathbf{h}_z(\boldsymbol{\theta}_s) - \boldsymbol{\mu}_t)^T, \tag{7}$$

where $\boldsymbol{\mu}_t$ is the local mean of the feature vectors in the interval.

We define a nonsymmetric $\boldsymbol{\Sigma}$-dependent squared distance between pairs of feature vectors as

$$a_{\boldsymbol{\Sigma}}^2(\mathbf{z}_t, \mathbf{z}_s) = (\mathbf{h}_z(\boldsymbol{\theta}_t) - \mathbf{h}_z(\boldsymbol{\theta}_s))^T \boldsymbol{\Sigma}_s^{-1} (\mathbf{h}_z(\boldsymbol{\theta}_t) - \mathbf{h}_z(\boldsymbol{\theta}_s)) \tag{8}$$

and the corresponding symmetric distance as

$$d_{\boldsymbol{\Sigma}}^2(\mathbf{z}_t, \mathbf{z}_s) = \frac{1}{2} \left(a_{\boldsymbol{\Sigma}}^2(\mathbf{z}_t, \mathbf{z}_s) + a_{\boldsymbol{\Sigma}}^2(\mathbf{z}_s, \mathbf{z}_t) \right). \tag{9}$$

Eq. (9) defines the Mahalanobis distance between the feature vectors. Thus, the distance in Eq. (9) is invariant under linear transformations, and by Lemma 1 and Eq. (6) it is invariant to the statistics of the measurement noise and to the corruption form (additive,

multiplicative, etc.). In addition, by Theorem 3.2 in ref. [14] and by Corollary 1, Eq. (9) approximates the Euclidean distance between the underlying controlling factors, i.e.,

$$\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_s\|^2 \approx d^2_{\boldsymbol{\Sigma}}(\mathbf{z}_t, \mathbf{z}_s) \tag{10}$$

by local linearization of the nonlinear transformation. For more details see refs. [7] and [14]. The dynamic described by independent Itô processes helps to identify a local "cloud" of each feature vector as a local time interval, which enables to compute the local covariance matrix in Eq. (7). In case the local clouds are provided, the computation of the noise-robust distance is independent of the dynamical model of the factors.

We note that the approximation in Eq. (10) is valid as long as the statistics of the noise are locally fixed (i.e., slowly changing compared to the fast variations of the controlling factors) and the fast variations of the controlling factors are emerged in the local histograms. The linear transformation employed by the measurement noise on the histogram may degrade the available information. For example, an additive Gaussian noise employs a low-pass "blurring" filter on the clean observation component. In case the dependency on the controlling factors is manifested in high-frequencies, the linear transformation employed by the noise significantly attenuates the connection between the measurements and the controlling factors. In practice, when the rank of the local covariance matrices of the feature vectors is lower than the number of the independent factors, it indicates that the available measurements are insufficient.

4.3 Anisotropic Kernel and Embedding

Let $\{\mathbf{z}_t\}_{t=1}^N$ be a finite interval of initially available measurements. We refer to these measurements as a reference set which is utilized to learn the model of the underlying controlling factors. Given the reference set, we construct an $N \times N$ nonsymmetric affinity matrix $\mathbf{A}$, whose (t,s)th element is given by

$$\mathbf{A}^{ts} = \exp\left\{-\frac{a^2_{\boldsymbol{\Sigma}}(\mathbf{z}_t, \mathbf{z}_s)}{\varepsilon}\right\}, \tag{11}$$

where $\varepsilon > 0$ is the kernel scale which may be set according to refs. [15] and [16]. Let $\mathbf{D} = \mathrm{diag}(\mathbf{A}\mathbf{1})$ be a diagonal normalizing matrix, where $\mathbf{1}$ is a vector of ones, and let $\bar{\mathbf{A}} = \mathbf{D}^{-1}\mathbf{A}$ be the corresponding normalized affinity matrix. The normalized matrix can be viewed as a Markov transition probability matrix for a jump process over the measurements. According to the diffusion analysis in ref. [6], the discrete Markov process converges to the continuous diffusion process that reveals the underlying structure of the data. We then define an $N \times N$ symmetric matrix $\mathbf{W}$ as

$$\mathbf{W} = \bar{\mathbf{A}}^T \bar{\mathbf{A}}. \tag{12}$$

It is shown in ref. [14] that each element of the symmetric matrix is proportional to a Gaussian with the Mahalanobis distance defined in Eq. (9), i.e.

$$\mathbf{W}^{ts} \propto \exp\left\{-\frac{d^2_{\boldsymbol{\Sigma}}(\mathbf{z}_t, \mathbf{z}_s)}{\varepsilon}\right\}.$$

Thus, $\mathbf{W}$ measures the affinity between the measurements according to the distance between the corresponding underlying factors. In addition, it encodes the temporal dynamics by incorporating local covariances and it is resilient to measurement noises. We note that further normalization of $\mathbf{W}$ may be applied to better handle nonuniform distribution of the measurements [4] [6].

Next, we compute the eigenvalues $\{\lambda_i\}_{i=1}^N$ and eigenvectors $\{\boldsymbol{\varphi}_i\}_{i=1}^N$ of $\mathbf{W}$. By refs. [7] and [14], the eigenvectors give an approximate parametrization of the low-dimensional manifold of the controlling factors. Specifically, the leading eigenvectors recover d candidates of the controlling factors up to a monotonic scaling. Without loss of generality, we may write

$$\varphi_i^t = \varphi_i(\theta_t^i),\ i = 1, \ldots, d; t = 1, \ldots, N,$$

where $\varphi_i(\cdot)$ are monotonic functions. We define a d-dimensional representation of the reference set by the following embedding

$$\boldsymbol{\Phi}(\mathbf{z}_t) \triangleq \left[\varphi_1^t, \varphi_2^t, \ldots, \varphi_d^t\right],\ t = 1, \ldots, N. \tag{13}$$

Since each eigenvector is a *monotonic* function of a controlling factor, the embedded domain organizes the measurements according to the values of the underlying controlling factors.

The eigenvectors of the kernel form a learned model based on the reference set. In addition, we store the feature vectors of the reference set along with their local covariance matrices to enable extension. The construction of the kernel in Eq. (12) yields that the eigenvectors of $\mathbf{W}$ are the singular right vectors of $\bar{\mathbf{A}}$. Thus, as discussed in ref. [14], the left singular vectors of $\bar{\mathbf{A}}$ naturally extend the spectral representation to any new measurement. Let $\mathbf{z}_t; t > N$ be a new measurement. Then, the spectral representation of the kernel can be efficiently extended by

$$\psi_i(\mathbf{z}_t) = \frac{1}{\lambda_i D^{ii}} \sum_{s=1}^{N} \exp\left\{-\frac{a_{\boldsymbol{\Sigma}}^2(\mathbf{z_t}, \mathbf{z}_s)}{\varepsilon}\right\} \varphi_i^s, \tag{14}$$

which involves only the stored information of the reference measurements. We define a d-dimensional representation of any measurement similarly to (13) by

$$\boldsymbol{\Psi}(\mathbf{z}_t) \triangleq \left[\psi_1(\mathbf{z}_t), \psi_2(\mathbf{z}_t), \ldots, \psi_d(\mathbf{z}_t)\right],\ t > N. \tag{15}$$

We note that the embedding does not take explicitly into account the dynamical model of the controlling factors. However, the Mahalanobis distance encodes the time dependency by using local covariance matrices, and the diffusion kernel reveals the dynamic by integrating those distances between the entire reference set.

The construction of $\mathbf{A}$ based on the Mahalanobis distance in Eqs. (8) and (11) entails a multivariate Gaussian mixture model in the histogram domain. The mixture consists of N infinitesimal Gaussian components, where each component s is centered at the reference histogram $\mathbf{h}_z(\boldsymbol{\theta}_s)$ and has a covariance $\boldsymbol{\Sigma}_s$. Accordingly, for any measurement at t, $\bar{\mathbf{A}}^{ts}$ can be viewed as the conditional probability that $\mathbf{h}_z(\boldsymbol{\theta}_t)$ is associated with with the sth component. It is further shown in ref. [17] that the extended embedding of a new measurement in Eq. (14) is optimal under the minimum mean square error (MMSE) criterion, i.e., $\psi_i(\mathbf{z}_t) = \frac{1}{\lambda_i}\mathbb{E}[\varphi_i^t|\mathbf{z}_t]$.

5 Bayesian Tracking

A nonparametric Bayesian framework to sequentially track the controlling factors is proposed by relying on the distribution in the embedded domain. In the Bayesian approach the posterior pdf of the factors is constructed based on the available information, including the measurements. Since the proposed embedding organizes the measurements according to the values of the controlling factors, we assume that the local distribution in the embedded domain is a good approximation to the local distribution in the controlling factors original domain. Thus, the desired posterior pdf can be estimated based on statistical information drawn from the density of the recovered factors. Recently, a Bayesian tracking approach in the embedded domain was proposed in ref. [18]. However, their embedding does not encode the temporal dependency and hence the dynamical model has to be adapted explicitly by assuming a flexible model.

In the spirit of Monte Carlo simulation methods and Gibbs sampling techniques we represent the posterior pdf by a set of samples. Let $\{\boldsymbol{\theta}_t^{(k)}(\boldsymbol{\theta}_{t-1}, \mathbf{z}_t)\}_{k=1}^P$ be a set of P support samples ("particles") that characterizes the posterior pdf at time t given the the factors at $t-1$ and the new measurement at t. Let $\{w_t^{(k)}(\boldsymbol{\theta}_{t-1}, \mathbf{z}_t)\}_{k=1}^P$ be a set of weights associated with the particles. For simplicity, in the remainder of the paper we omit the dependency on $\boldsymbol{\theta}_{t-1}$ and $\mathbf{z}_t$. The posterior pdf at t can be approximated as

$$p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \mathbf{z}_t) \approx \sum_{i=1}^{P} w_t^{(k)} \delta\left(\boldsymbol{\theta}_t - \boldsymbol{\theta}_t^{(k)}\right), \tag{16}$$

where the weights are denoted by

$$w_t^{(k)} \triangleq p(\boldsymbol{\theta}_t^{(k)}|\boldsymbol{\theta}_{t-1}, \mathbf{z}_t)$$

with $\sum_{k=1}^P w_t^{(k)} = 1$. We therefore have a discrete weighted approximation of the desired posterior pdf. By Bayes theorem and the Markov dynamical model we obtain

$$w_t^{(k)} \propto p(\boldsymbol{\theta}_t^{(k)}|\boldsymbol{\theta}_{t-1}) p(\mathbf{z}_t|\boldsymbol{\theta}_t^{(k)}). \tag{17}$$

The densities in Eq. (17) are estimated based on the embedded domain. Quantities associated with the embedding are denoted with tilde, e.g., $\tilde{\boldsymbol{\theta}}_t$ denotes the embedding associated with $\boldsymbol{\theta}_t$. The likelihood function is determined by

$$\begin{aligned} &p(\mathbf{z}_t|\boldsymbol{\theta}_t^{(k)}) \propto \\ &\exp\left\{-\left(\boldsymbol{\Psi}(\mathbf{z}_t) - \tilde{\boldsymbol{\theta}}_t^{(k)}\right)^T \left[\tilde{\boldsymbol{\Sigma}}_t^{(k)}\right]^{-1} \left(\boldsymbol{\Psi}(\mathbf{z}_t) - \tilde{\boldsymbol{\theta}}_t^{(k)}\right)\right\}, \end{aligned} \tag{18}$$

where $\tilde{\boldsymbol{\Sigma}}_t^{(k)}$ is the local covariance of embedded measurements near $\tilde{\boldsymbol{\theta}}_t^{(k)}$. Let $\tilde{\mathcal{N}}_{t-1}$ be a set of time indices of samples in a $\xi > 0$ neighborhood of $\tilde{\boldsymbol{\theta}}_{t-1}$, defined as

$$\tilde{\mathcal{N}}_{t-1} = \left\{ s \left| \left\| \tilde{\boldsymbol{\theta}}_s - \tilde{\boldsymbol{\theta}}_{t-1} \right\| < \xi, s < t-1 \right. \right\}.$$

The samples in the neighborhood have known dynamics since their succeeding samples are available. Accordingly, we collect the set of succeeding samples, i.e., $\left\{\tilde{\boldsymbol{\theta}}_{s+1}\right\}_s$ for $s \in \tilde{\mathcal{N}}_{t-1}$,

and compute its mean and its covariance, denoted by $\tilde{\boldsymbol{\theta}}_{\tilde{\mathcal{N}}_{t-1}}$ and $\tilde{\boldsymbol{\Sigma}}_{\tilde{\mathcal{N}}_{t-1}}$, respectively. Then, the pdf of the dynamic of the factors is estimated by

$$p(\boldsymbol{\theta}_t^{(k)}|\boldsymbol{\theta}_{t-1}) \propto \exp\left\{-\left(\tilde{\boldsymbol{\theta}}_t^{(k)} - \tilde{\boldsymbol{\theta}}_{\tilde{\mathcal{N}}_{t-1}}\right)^T \left[\tilde{\boldsymbol{\Sigma}}_{\tilde{\mathcal{N}}_{t-1}}\right]^{-1} \left(\tilde{\boldsymbol{\theta}}_t^{(k)} - \tilde{\boldsymbol{\theta}}_{\tilde{\mathcal{N}}_{t-1}}\right)\right\}. \tag{19}$$

The estimation of the likelihood function in Eq. (18) implies a local multivariate Gaussian distribution in the embedded domain. The usage of local densities may overcome the nonuniform density in the embedded domain near the boundaries. The embedding of a new measurement at t, given that it is associated with the kth particle, is a normal random vector with mean $\tilde{\boldsymbol{\theta}}_t^{(k)}$ and covariance $\tilde{\boldsymbol{\Sigma}}_t^{(k)}$. Thus, Eq. (18) measures the likelihood that the embedding of the measurement at t is associated with the kth particle. Similarly, the estimation of the pdf of the dynamic in Eq. (20) implies that the embedding of the kth particle, given the factors at $t-1$, is a normal random vector with mean $\tilde{\boldsymbol{\theta}}_{\tilde{\mathcal{N}}_{t-1}}$ and covariance $\tilde{\boldsymbol{\Sigma}}_{\tilde{\mathcal{N}}_{t-1}}$. Thus, Eq. (20) measures the probability that the kth particle is on the trajectory that passes through $\tilde{\boldsymbol{\theta}}_{t-1}$ at $t-1$. We note that the estimation of the dynamic is similar in spirit to the prediction step in Kalman filtering. Instead of using a predefined dynamical model, we estimate the empirical transitions in each step independently.

Based on the estimate of the posterior pdf in Eq. (16), a new set of particles can be drawn

$$\boldsymbol{\theta}_{t+1}^{(k)} \sim p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \mathbf{z}_t).$$

Since it might be difficult to draw samples from the posterior pdf, we use the assumption of a multivariate Gaussian model and draw samples from normal distributions. In practice, in case a sufficient number of measurements are available, the new particles can simply be the neighbors in the embedded space. According to Eqs. (18) and (20), the particles are the embedding of factors at s such that $s < t$ that lie near the embedding of the measurement at t ("likelihood") and near the embedding of the succeeding factors of similar trajectories from previous steps ("dynamic").

Based on the estimate of the posterior pdf, an MMSE estimator of the factors at t can be computed by

$$\begin{aligned}\hat{\boldsymbol{\theta}}_t &= \mathbb{E}\left[\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \mathbf{z}_t\right] \\ &= \int \boldsymbol{\theta}_t p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \mathbf{z}_t) d\boldsymbol{\theta}_t \\ &= \sum_{i=1}^{P} p(\boldsymbol{\theta}_t^{(k)}|\boldsymbol{\theta}_{t-1}, \mathbf{z}_t)\boldsymbol{\theta}_t^{(k)}. \end{aligned} \tag{20}$$

In case the particles are nearest neighbors, then their estimates are already computed in previous time steps. However, for initial estimates in Eq. (20), we require the values of the original factors to align the embedding of each particle $\tilde{\boldsymbol{\theta}}_t^{(k)}$ with the actual value $\boldsymbol{\theta}_t^{(k)}$.

An interesting comment is that the likelihood function based on the embedded domain provides a foundation for synthesis and tracking in the measurement domain. This ability is highly beneficial in a wide variety of applications and will be addressed in future work.

6 Application of Stochastic Sensing to the Toy Example

We simulate 2-dimensional trajectories of the two independent controlling factors according to Eq. (4) and the corresponding noisy measurements under the three measurement schemes. A short segment of a 2-dimensional trajectory is depicted in Fig. 1 and a corresponding trajectory on the sphere is presented in Fig. 2.

The first $N = 2000$ samples of measurements are used as the reference set, which empirically was shown to be a sufficient amount of data to represent the dynamical model. For each reference measurement we compute a histogram in a short window, which approximates the temporal pdf of the measurement, and a local covariance matrix according to Eq. (7). Next, the kernels are calculated by Eqs. (11) and (12). Finally, the embedding of the reference measurements is constructed based on the spectral decomposition of the kernels by Eq. (13).

Measurements at times $t > 2000$ are sequently processed. For each measurement the extended embedding in Eq. (15) is computed. Then, the temporal posterior pdf is estimated according to the proposed Bayesian framework by relying on the embedded domain. Since a sufficient amount of "historic" data is available, the particles are chosen as nearest neighbors in the embedded domain and not drawn from the posterior distribution. Finally, the MMSE estimator of the factors at each time step t is calculated according to Eq. (20).

Figure 3 depicts three eigenvectors that correspond to the same movement of the object and represent the vertical angle. Each eigenvector is obtained under a different measurement scheme. We note that the presented 2000 coordinates of the eigenvectors are computed by extension. The measurements under the three schemes are very different in their nature, e.g., spike sequences in Schemes 1 and 2 compared to noisy 3-dimensional coordinates in Scheme 3. We obtain similar eigenvectors which implies intrinsic modeling and demonstrates the invariability of the proposed approach to the type of the acquisition system and to measurement noises. In order to further demonstrate resilience to measurement noise, we present in Fig. 4 three eigenvectors obtained under Scheme 1 with three different noise levels.

Figure 5 presents the Bayesian tracking of the vertical angle of the movement under Scheme 1 in a time interval following the reference segment. The yellow line is the trajectory of the vertical angle. The vertical gray level at each time slot represents the posterior pdf estimate. The dotted black line is the expected value based on the posterior pdf estimate.

The proposed stochastic sensing approach is compared to the SIR particle filter [13] in order to provide objective evaluation. This particular sequential Monte Carlo method was chosen due to its simplicity. The particle filter requires the dynamical and measurement models along with the noise statistics as priors, whereas the stochastic sensing approach utilizes historic reference data. The results under measurement Scheme 1 with different noise levels are summarized in Table 1. The results are averaged over several realizations of measurements and noise and several repetitions of the stochastic particle filter tracking. We note that the typical spike rate ranges between 5 and 15 and the measurement noise rate ranges between 0 and 5. Thus, we used 20 histogram bins equally divided in $[0, 20]$. We observe that the stochastic sensing approach outperforms the particle filter. The proposed approach exploits the new measurement by embedding it into a domain that implicitly encodes the dynamical model of the data. On the other hand, the SIR particle filter uses an

importance sampling measure that ignores the new measurement. In addition, the performance of the particle filter degrades as the noise level increases, whereas the performance of stochastic sensing is invariance to noise levels and depends on the compatibility of the measurement values to the choice of predefined histogram bins.

7 Acoustic Localization

We revisit the experiment presented in [19] and test the ability of the proposed algorithm to recover the location of a sound source. Following is a brief description of the experiment setup. Inside a reverberant room, we positioned an omni-microphone in a fixed position. A long "arm" was connected to the base of the microphone, and attached to a turntable that controlled the horizontal angle of the arm. A speaker was located on the far-end of the arm. Thus, the turntable controlled the direction of arrival of the sound played by the speaker with respect to the microphone. Using the turntable we tested 60 different horizontal angles with 1° spacing. From each angle, a zero-mean and unit variance white Gaussian noise was played from the sound source and recorded in the microphone. The movement of the arm along the entire range of 60 angles was repeated several times.

We note that the horizontal angle constitutes the sole degree of freedom in this experiment, as the rest of the room parameters are fixed. Mild variations caused by the movement of the arm are neglected. We calculate histograms of the recorded raw signal in short-time windows and compute their corresponding local covariance matrices. Then we build the kernel and obtain the embedding. Similar to the results reported in [19], the leading eigenvector represents the horizontal angle. The recovery of the angle based on the embedding yields a mean square error of 2°. This result is inferior to the result obtained in [19]. However in [19], a-priori knowledge of noiseless Gaussian recordings was utilized. In addition, the features were computed is several scales. On the other hand, no prior knowledge is used in our approach and the short-term histograms are computed on the raw data.

References

[1] J. Tenenbaum, V. de Silva, and J. Langford, *A global geometric framework for nonlinear dimensionality reduction*, Science, 290 (2000), pp. 2319–2323.

[2] D. Donoho and C. Grimes, *Hessian eigenmaps: new locally linear embedding techniques for high-dimensional data*, Proc. Nat. Acad. Sci., 100 (2003), pp. 5591–5596.

[3] M. Belkin, and P. Niyogi, *Laplacian Eigenmaps for dimensionality reduction and data representation*, Neural Computation, 15 (2003), pp. 1373–1396.

[4] R.R. Coifman, S. Lafon, A. Lee, M. Maggioni, B. Nadler, F. Warner, and S. Zucker, *Geometric diffusions as a tool for harmonic analysis and structure definition of data. Part I: Diffusion maps*, Proc. of Nat. Acad. Sci., (2005), pp. 7426–7431.

[5] R.R. Coifman, S. Lafon, A. Lee, M. Maggioni, B. Nadler, F. Warner, and S. Zucker, *Geometric diffusions as a tool for harmonic analysis and structure definition of data. Part II: Multiscale methods*, Proc. of Nat. Acad. Sci., (2005), pp. 7432–7437.

[6] R.R. Coifman and S. Lafon, *Diffusion maps*, Appl. Comp. Harm. Anal., 21 (2006), pp. 5–30.

[7] A. Singer, and R.R. Coifman, *Non-linear independent component analysis with diffusion maps*, Appl. Comp. Harm. Anal., 25 (2008), pp. 226–239.

[8] R. Talmon, D. Kushnir, R.R. Coifman, I. Cohen, and S. Gannot, *Parametrization of linear systems using diffusion kernels*, IEEE Trans. Signal Process., 60 (2012).

[9] Y. Bar-Shalom, *Tracking and data association*, Academic Press Professional, (1987).

[10] R.E. Kalman, *A new approach to linear filtering and prediction problems*, Trans. ASME J. Basic Eng., 82 (1960), pp. 34-45.

[11] S.J. Julier, and J.K. Uhlmann, *Unscented filtering and nonlinear estimation*, Proc. of the IEEE, 92 (2004), pp. 401–422.

[12] A. Doucet, S. Godsill, and C. Andrieu, *On sequential Monte Carlo sampling methods for Bayesian filtering*, Statistics and Computing, 10 (2000), pp. 197–208.

[13] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, *A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking*, IEEE Trans. Signal Process., 50 (2003), pp. 174–188.

[14] D. Kushnir, A. Haddad, and R.R. Coifman, *Anisotropic diffusion on sub-manifolds with application to earth structure classification*, Appl. Comp. Harm. Anal., 32 (2012), pp. 280–294.

[15] M. Hein, and Y. Audibert, *Intrinsic dimensionality estimation of submanifolds in* $\mathbb{R}^d$, In Proceedings of the 22nd International Conference on Machine Learning, eds. L. De Raedt, S. Wrobel (Association for Computing Machinery, New York), (2005), pp. 289–296.

[16] A. Singer, R. Erban, I.G. Kevrekidis, and R.R. Coifman, *Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps*, Proc. of Nat. Acad. Sci., (2009), pp. 16090–16095.

[17] R. Talmon, I. Cohen, S. Gannot, and R.R. Coifman, *Supervised graph-based processing for sequential transient interference suppression*, submitted to IEEE Trans. Audio Speech Lang. Process., (2012).

[18] H. Chen, E. Fox, J. Silva, D. Dunson, and L. Carin, *Hierarchical bayesian embeddings for analysis and synthesis of dynamic data*, submitted, (2012).

[19] R. Talmon, I. Cohen, and S. Gannot, *Supervised source localization using diffusion kernels*, Proc. IEEE Work. Appl. Sig. Process. Audio Acous., (2011).

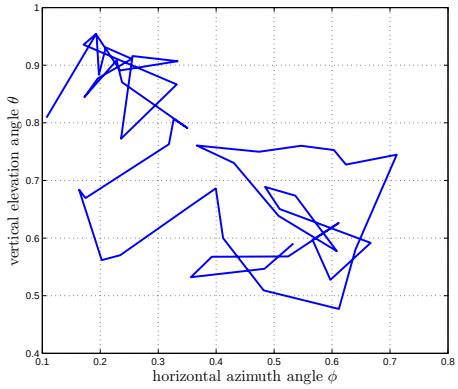

Figure 1: A segment of the 2-dimensional trajectory of the two independent factors: the horizontal and vertical angles.

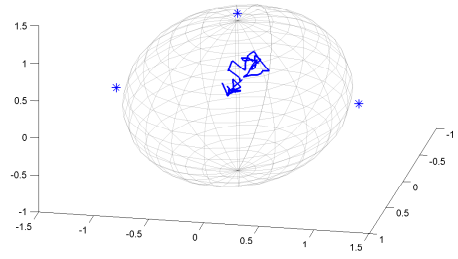

Figure 2: A segment of the 3-dimensional movement of the object on the sphere. The locations of the 3 sensors are marked with $*$.

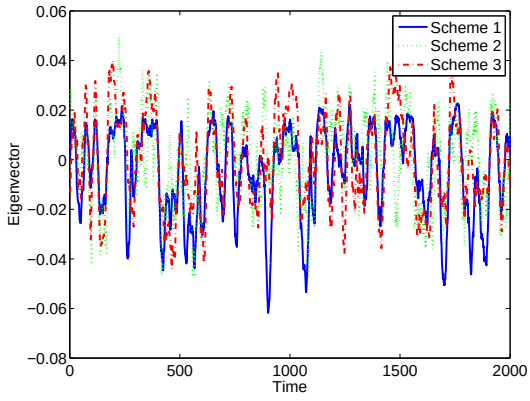

Figure 3: A comparison between the obtained eigenvectors (corresponding to the vertical angle) under the three different measurement schemes (measuring the same movement).

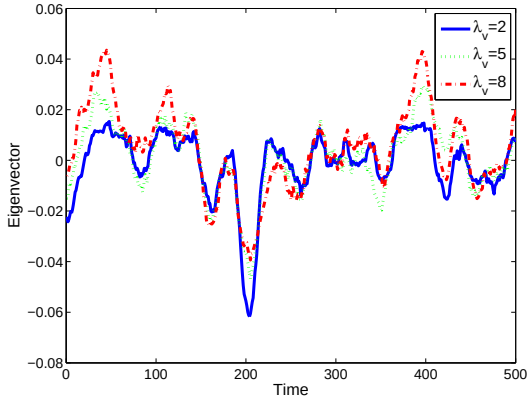

Figure 4: A comparison between the obtained eigenvectors (corresponding to the vertical angle) under the first measurement scheme with different noise levels ($\lambda_v = 2, 5, 8$).

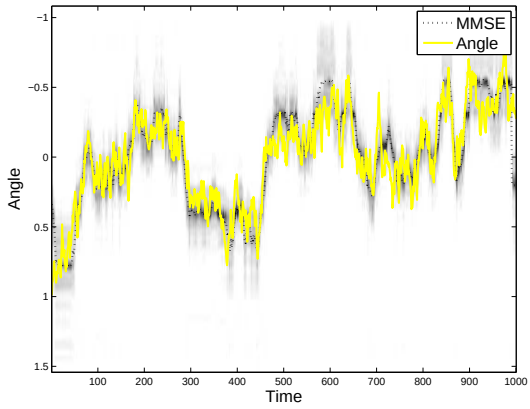

Figure 5: Tracking the horizontal angle of the movement in a time interval following the reference segment. The yellow line is the trajectory of the horizontal angle. The gray level represents the posterior pdf estimate obtained by the Bayesian tracking. The dotted black line is the expected value based on the posterior pdf estimate (MMSE estimator).

Table 1: The MSE obtained under Scheme 1 in dB.

	Stochastic Sensing		SIR Particle Filter	
Noise Rate	Horizontal Angle	Vertical Angle	Horizontal Angle	Vertical Angle
$\lambda_v = 0$	-5.3	-3.7	-3.1	-2.7
$\lambda_v = 1$	-6.5	-4.1	-3.0	-2.5
$\lambda_v = 2$	-6.2	-4.4	-2.9	-2.4
$\lambda_v = 3$	-5.9	-4.6	-2.5	-2.3
$\lambda_v = 4$	-4.9	-4.7	-2.4	-2.2
$\lambda_v = 5$	-4.1	-4.8	-2.0	-1.9