Empirical intrinsic geometry for nonlinear modeling and time series filtering

Ronen Talmon[1] and Ronald R. Coifman[1]

Department of Mathematics, Yale University, New Haven, CT 06520

Contributed by Ronald R. Coifman, April 18, 2013 (sent for review February 14, 2012)

In this paper, we present a method for time series analysis based on empirical intrinsic geometry (EIG). EIG enables one to reveal the low-dimensional parametric manifold as well as to infer the underlying dynamics of high-dimensional time series. By incorporating concepts of information geometry, this method extends existing geometric analysis tools to support stochastic settings and parametrizes the geometry of empirical distributions. However, the statistical models are not required as priors; hence, EIG may be applied to a wide range of real signals without existing definitive models. We show that the inferred model is noise-resilient and invariant under different observation and instrumental modalities. In addition, we show that it can be extended efficiently to newly acquired measurements in a sequential manner. These two advantages enable us to revisit the Bayesian approach and incorporate empirical dynamics and intrinsic geometry into a nonlinear filtering framework. We show applications to nonlinear and non-Gaussian tracking problems as well as to acoustic signal localization.

anisotropic diffusion | manifold learning | signal processing

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 (1–4). The method proposed by Singer and Coifman (5) is of particular interest because the data are assumed to be inaccessible and can be observed only via unknown nonlinear functions. By integrating local principal components with diffusion maps, the approach of Singer and Coifman (5) provides modeling of the underlying parametric manifold, whereas classical manifold learning methods provide a parameterization of the observable manifold. However, geometric methods usually suffer from two significant disadvantages. First, in many natural systems, the mapping of the low-dimensional data into a subset of high-dimensional observations is stochastic. In addition, a random measurement noise usually corrupts the observations. Thus, the geometry of the observations may not convey the appropriate information on the underlying parametric manifold. Second, these methods provide modeling to a given dataset and do not model a stream of new incoming measurements well.

In this paper, we propose a framework for sequential processing of time series based on empirical intrinsic geometric models. We adopt the nonlinear filtering formalism and propose a method consisting of two steps: novel data-driven modeling of stochastic data, which is resilient to noise as well as invariant to the measurement modality, and Bayesian filtering of new incoming measurements based on the learned model. High-dimensional time series often exhibit highly redundant representations and can be compactly represented by a dynamical process on a low-dimensional manifold. Thus, in the first stage, we reveal the low-dimensional manifold and infer the underlying process using anisotropic diffusion. The proposed approach relies on the observation that the local statistics of the measurements convey the geometric information on the underlying factors rather than the specific realization at hand (6, 7). Thus, we empirically learn the underlying manifold of local probability densities and recover their dynamics. In the second stage, we define a Bayesian framework in which the likelihood function is defined based on the learned model and the dynamics are obtained empirically from past measurements. In particular, we extend the learned model to a stream of incoming new measurements and track the underlying process on the low-dimensional manifold.

The presented framework may overcome a major limitation of classical nonlinear filtering methods (8, 9). We describe a nonparametric method that implicitly infers the dynamical and measurement models, whereas existing methods require the dynamical and measurement models as well as the noise statistics as priors. This ability is critical for processing real signals in which the underlying process is unknown (e.g., financial time series, biomedical signals), and hence no samples of the underlying process are available to train a predefined model.

Dynamical and Measurement Models

Let $\mathbf{y}_t$ denote an n-dimensional observation process in time index t, drawn from a conditional time-varying probability density function (pdf). Consider a model in which the pdf, denoted by $f(\mathbf{y};\boldsymbol{\theta})$, is controlled by a d-dimensional process $\boldsymbol{\theta}_t$ consisting of underlying factors. The dynamics of the underlying process are described by normalized stochastic differential equations

$$d\theta_t^i = a^i\left(\theta_t^i\right)dt + dw_t^i, \quad i = 1,\ldots,d, \qquad [1]$$

where a^i are unknown drift coefficients and $\dot{w}_t^i$ are independent white noises. In this paper, subscripts represent labeling and superscripts represent access to elements in vectors or matrices.

The observation process $\mathbf{y}_t$ is accessible only via an n-dimensional noisy measurement process $\mathbf{z}_t$, given by

$$\mathbf{z}_t = g(\mathbf{y}_t, \mathbf{v}_t), \qquad [2]$$

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

Our goal is twofold: (1) recovering $\boldsymbol{\theta}_t$ and building a dynamical model on the underlying manifold. Based on a finite set of noisy measurements, we infer an empirical intrinsic geometric model, (2) extending the obtained model when new measurements $\mathbf{z}_t$ become available and tracking the underlying process $\boldsymbol{\theta}_t$. We note that this general state-space formalism of the problem, adopted from nonlinear filtering studies (10), may apply to many real-world applications.

The construction of the model will be carried out in three time scales. The *micro* time scale is the domain of the measurements $\mathbf{z}_t$, where we assume that the sampling rate is sufficiently high to enable us to estimate the time-varying densities of the measurements in short windows. The *mezzo* scale is the domain of the conditional densities $p(\mathbf{z};\boldsymbol{\theta})$. In this domain, the densities are assumed to be

Author contributions: R.T. and R.R.C. designed research; R.T. and R.R.C. performed research; R.T. and R.R.C. contributed new reagents/analytic tools; R.T. and R.R.C. analyzed data; and R.T. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. E-mail: ronen.talmon@yale.edu or coifman@math.yale.edu.

slowly changing and we are able to estimate their local variations. Finally, the *macro* scale will be the anisotropic diffusion domain, where all the available information is aggregated into a global model. These implicit assumptions imply that the obtained model captures the slow factors in the data. Such a result was previously demonstrated on chemical reactions by Singer et al. (11) and extensively investigated in slow feature analysis studies (12).

Toy Example

We track a radiating object moving on a 3D sphere. Because 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:

$$\dot{\theta}_t^i = -\nabla U(\boldsymbol{\theta}_t) + \dot{w}_t^i, \quad i = 1, 2, \qquad [3]$$

where $\dot{w}_t^i$ are independent white Gaussian noises and U is a potential field. Let $\mathbf{x}_t$ denote the 3D coordinates of the object position on the sphere. We consider three measurement modalities, which include both continuous and point process domains, to demonstrate the universality and robustness of our method. In modality 1, the radiation of the object is measured by three "Geiger counter" sensors positioned at $\mathbf{s}_j, j = 1, 2, 3$ outside the sphere (the setup is illustrated in Fig. 1). The sensors detect the object and fire spikes through a spatial point process at a varying rate that depends on the proximity of the object to the sensors. The spikes are fired by each sensor according to a Poisson distribution with rate $\lambda^j(\boldsymbol{\theta}_t) = \exp\{-\|\mathbf{s}_j - \mathbf{x}_t\|\}$. 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

$$z_t^j = g^j(y_t, v_t) = y_t^j + v_t^j, \quad j = 1, 2, 3,$$

where v_t^j is a spike train drawn from a Poisson distribution with a fixed rate λ_v^j. Modality 2 is similar to modality 1. Each sensor fires spikes randomly according to the proximity of the object. However, in this modality, we simulate sensors with unreliable clocks. We measure the time interval between two consecutive spikes, which is given by $\mathbf{z}_t = \mathbf{y}_t + \mathbf{v}_t$. Suppose y_t^j is drawn from an exponential distribution with a rate parameter $\lambda^j(\boldsymbol{\theta}_t)$, and suppose v_t^j is drawn from a fixed normal distribution representing the clock inaccuracy. In modality 3, we consider a measurement in a continuous domain. We use three sensors that measure the position of the source directly, that is, $\mathbf{z}_t = \mathbf{x}_t + \mathbf{v}_t$, where $\mathbf{v}_t$ is an additive Gaussian white noise.

In this example, the goal is to reveal and track the 2D trajectory $\boldsymbol{\theta}_t$ of the horizontal and vertical angles sequentially based on a stream of incoming noisy measurements $\mathbf{z}_t$. This problem exhibits a nonlinear and non-Gaussian challenging setting, and it has been inspired by a variety of nonlinear filtering studies (e.g., refs. 13, 14).

Intrinsic Modeling

Local Densities and Histograms. Lemma 1. *The pdf of the measured process* $\mathbf{z}_t$, *i.e.*, $p(\mathbf{z};\boldsymbol{\theta})$, *is given by a linear transformation of the pdf of the clean observation component* $\mathbf{y}_t$, *i.e.*, $f(\mathbf{y};\boldsymbol{\theta})$.

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

$$p(\mathbf{z};\boldsymbol{\theta}) = \int_{g(\mathbf{y},\mathbf{v})=\mathbf{z}} f(\mathbf{y};\boldsymbol{\theta}) q(\mathbf{v}) d\mathbf{y} d\mathbf{v}. \qquad [4]$$

In practice, the time-varying pdfs $p(\mathbf{z};\boldsymbol{\theta})$ need to be estimated. Let $\mathbf{h}_t$ be an m-bins histogram of the measurements in a short window of length N_h centered at time t, whose jth element is approximated by

$$h_t^j = \int_{\mathbf{z}\in\mathcal{H}_j} p(\mathbf{z};\boldsymbol{\theta}) d\mathbf{z}, \qquad [5]$$

where $\mathcal{H}_j$ is the jth bin. By *Lemma 1* and by Eq. **5**, we get the following result.

Corollary 1. *The empirical distribution* $\mathbf{h}_t$ *is given by a linear transformation of the pdf of the clean measurement component* $\mathbf{y}_t$.

In other words, any measurement noise is expressed as a linear transformation in the histograms domain.

Noise-robust Metric. We view each histogram $\mathbf{h}_t$ as a feature vector for each measurement $\mathbf{z}_t$. For each feature vector $\mathbf{h}_t$, we estimate the local covariance matrix $\mathbf{C}_t$ in a time window of length $2N_c + 1$ centered at t according to

$$\hat{\mathbf{C}}_t = \frac{1}{2N_c + 1} \sum_{\tau=t-N_c}^{t+N_c} (\mathbf{h}_\tau - \hat{\boldsymbol{\mu}}_t)(\mathbf{h}_\tau - \hat{\boldsymbol{\mu}}_t)^T, \qquad [6]$$

where $\hat{\boldsymbol{\mu}}_t$ is the empirical mean of the feature vectors in the window.

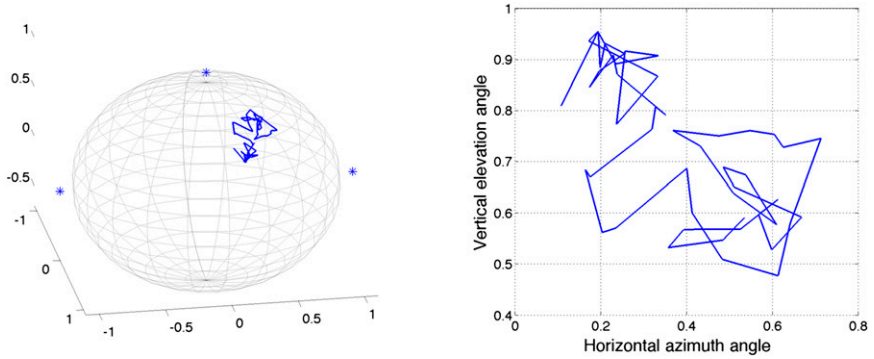

Fig. 1. Toy example setup. (*Left*) Segment of the 3D movement of the object on the sphere. The locations of the three sensors are marked with an asterisk. (*Right*) Corresponding segment of the 2D trajectory of the two independent factors: the horizontal and vertical angles.

Using the feature vectors and their corresponding local covariance matrices, we define the Mahalanobis distance between the measurements as

$$d^2(\mathbf{z}_t,\mathbf{z}_\tau)=(\mathbf{h}_t-\mathbf{h}_\tau)^T(\mathbf{C}_t+\mathbf{C}_\tau)^{-1}(\mathbf{h}_t-\mathbf{h}_\tau). \quad [7]$$

Because the Mahalanobis distance is invariant under linear transformations of $\mathbf{h}_t$, by *Corollary 1*, it is invariant to any measurement noise. Using Itô calculus, it was shown by Singer and Coifman (5) that $\mathbf{C}_t=\mathbf{J}_t\mathbf{J}_t^T$, where $\mathbf{J}_t$ is the Jacobian of the function $\mathbf{h}_t=h(\boldsymbol{\theta}_t)$. By assuming that h is a bi-Lipschitz function of the underlying process and using local linearization of the function (i.e., $\mathbf{h}_t=\mathbf{J}_t\boldsymbol{\theta}_t+\epsilon_t$), it can be further shown that the Mahalanobis distance approximates the Euclidean distance between the corresponding samples of the underlying process to a second order, that is,

$$\|\boldsymbol{\theta}_t-\boldsymbol{\theta}_\tau\|^2=d^2(\mathbf{z}_t,\mathbf{z}_\tau)+\mathcal{O}\left(\|\mathbf{h}_t-\mathbf{h}_\tau\|^4\right). \quad [8]$$

We note that the bi-Lipschitz assumption describes the limitations of the aforementioned result. On one hand, the function h must be sufficiently smooth to describe the underlying process accurately. On the other hand, the function h must be sensitive to small variations of the underlying process. This guarantees that the local covariance matrices capture the variability of the underlying process and enables one to obtain the approximation in Eq. **8**. For example, a high-variance additive Gaussian noise might obscure the connection between the measurements and the underlying processes. In such cases, $p(\mathbf{z};\boldsymbol{\theta})$, and hence h, becomes a degenerate function of $\boldsymbol{\theta}$, which is not bi-Lipschitz.

We are able to show that in case the feature vectors are defined as the logarithm of the histograms, the local covariance matrix has a tight connection to the Fisher information matrix. Furthermore, in this case, the proposed invariant metric is related to the Kullback–Liebler divergence and may define a dynamical process on the low-dimensional manifold (15). Details on the relationship between empirical intrinsic geometry (EIG) and information geometry are provided by Talmon and Coifman (16).

Anisotropic Kernel and Embedding. Let $\{\overline{\mathbf{z}}_s\}_{s=1}^N$ be an initially available interval of measurements. We refer to these measurements as a reference set $\mathcal{R}$, which is used to learn the model of the underlying process. Given the reference set, we construct an $N\times N$ affinity matrix, whose (s,s')-th element is given by

$$W_{\mathcal{R}}^{ss'}=\exp\left\{-\frac{d^2(\overline{\mathbf{z}}_s,\overline{\mathbf{z}}_{s'})}{\varepsilon}\right\}, \quad [9]$$

where $\varepsilon>0$ is the kernel scale, which may be set according to Singer et al. (11) and Hein et al. (17). Thus, $\mathbf{W}_{\mathcal{R}}$ measures the affinity between the measurements according to the distance between the corresponding samples of the underlying process. It is invariant to the observation modality, and it is resilient to measurement noise. Let $\mathbf{D}$ be a diagonal matrix whose diagonal elements are $D^{ss}=\sum_{s'}W_{\mathcal{R}}^{ss'}$. Let $\tilde{\mathbf{W}}_{\mathcal{R}}=\mathbf{D}^{-1/2}\mathbf{W}_{\mathcal{R}}\mathbf{D}^{-1/2}$ be a normalized kernel that shares its eigenvectors with the normalized graph-Laplacian $\mathbf{I}-\tilde{\mathbf{W}}_{\mathcal{R}}$ (18). We compute the eigenvalues and eigenvectors $\{\tilde{\lambda}_i,\tilde{\boldsymbol{\varphi}}_i\}$, $i=1,\ldots,N$ of $\tilde{\mathbf{W}}_{\mathcal{R}}$. By the methodology of Singer and Coifman (5). and Kushnir et al. (19), $\tilde{\mathbf{W}}_{\mathcal{R}}$ converges to a diffusion operator that reveals the low-dimensional manifold and the eigenvectors give a parameterization of the underlying processes. Specifically, in case of independent factors, the eigenvectors recover d proxies for the controlling factors. We assume that these proxies are the d principal eigenvectors associated with the largest eigenvalue, although there is no guarantee. Guidelines for choosing the eigenvectors are provided by Singer (20). We define a d-dimensional representation of each reference measurement by the following embedding

$$\boldsymbol{\Phi}(\overline{\mathbf{z}}_s)=\left[\tilde{\varphi}_1^s,\tilde{\varphi}_2^s,\ldots,\tilde{\varphi}_d^s\right],\quad s=1,\ldots,N. \quad [10]$$

The eigenvectors of the kernel form a learned model based on the reference set, and the embedded domain organizes the reference measurements according to the values of the underlying factors.

When new measurements $\mathbf{z}_t; t>N$ become available, we define a nonsymmetrical affinity matrix $\mathbf{A}$ between the new measurements and the reference measurements, whose (t,s)-th element is given by

$$A^{ts}=\exp\left\{-\frac{a^2(\mathbf{z}_t,\overline{\mathbf{z}}_s)}{\varepsilon}\right\}, \quad [11]$$

where $a^2(\mathbf{z}_t,\mathbf{z}_s)$ is a nonsymmetrical square distance, defined by

$$a^2(\mathbf{z}_t,\overline{\mathbf{z}}_s)=(\mathbf{h}_t-\mathbf{h}_s)^T\mathbf{C}_s^{-1}(\mathbf{h}_t-\mathbf{h}_s). \quad [12]$$

We note that the rank of the covariance matrices is d; hence, a pseudoinverse should be used. In practice, when the rank of the empirical covariance matrices is smaller than d, it indicates that the available measurements are insufficient. Let $\tilde{\mathbf{A}}=\mathbf{D}_A^{-1}\mathbf{A}\boldsymbol{\omega}^{-1}$, where $\mathbf{D}_A$ is a diagonal matrix whose diagonal elements are the sums of rows of $\mathbf{A}$ and $\boldsymbol{\omega}$ is a diagonal matrix whose diagonal elements are the sums of columns of $\mathbf{D}_A^{-1}\mathbf{A}$.

It is shown by Kushnir et al. (19) that $\mathbf{W}_{\mathcal{R}}=\tilde{\mathbf{A}}^T\tilde{\mathbf{A}}$. In addition, the kernel $\mathbf{W}=\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T$ can be viewed as an extended kernel, whose elements are shown by Talmon et al. (21) to be proportional to a Gaussian defined similar to Eq. **9**, with the Mahalanobis distance between the new measurements. The relationship between the kernels $\mathbf{W}_{\mathcal{R}}$ and $\mathbf{W}$ implies that (1) the kernels share the same eigenvalues λ_i; (2) the eigenvectors $\boldsymbol{\varphi}_i$ of $\mathbf{W}_{\mathcal{R}}$ are the *right* singular vectors of $\tilde{\mathbf{A}}$; and (3) the eigenvectors $\boldsymbol{\psi}_i$ of $\mathbf{W}$ are the *left* singular vectors of $\tilde{\mathbf{A}}$, which naturally extend the spectral representation to any new measurement. As a result, the definition of the singular value decomposition given by

$$\psi_i^t=\frac{1}{\sqrt{\lambda_i}}\sum_{s=1}^N\tilde{A}^{ts}\varphi_i^s, \quad [13]$$

describes an efficient extension of the spectral representation to a new measurement at time t without applying additional eigendecomposition. We note that Eq. **13** involves only the available information on the reference measurements. In particular, the computation does not require the local covariance matrix of the new measurement, which is adequate to real-time processing because it circumvents the lag required to collect a local neighborhood for each new measurement. Based on the extended eigenvectors, we define a d-dimensional representation of any measurement similar to Eq. **10** by

$$\boldsymbol{\Psi}(\mathbf{z}_t)=\left[\psi_1^t,\psi_2^t,\ldots,\psi_d^t\right],\ t>N. \quad [14]$$

The construction of $\mathbf{A}$ based on Eqs. **11** and **12** implies 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 $\mathbf{h}_s$ and has a covariance $\mathbf{C}_s$. For any measurement at t, A^{ts} can be viewed as the conditional probability that $\mathbf{h}_t$ is associated with the sth component and $W^{t\tau}$ can be viewed as the probability that $\mathbf{h}_t$ and $\mathbf{h}_\tau$ are associated with the same component. It is further shown by Talmon et al. (21) that the coordinates of the extended embedding

are proportional to the minimum mean square error (MMSE) estimator.

Bayesian Tracking

The presented embedding does not take explicitly into account the chronological order of the measurements. However, the Mahalanobis distance encodes the time dependency by using local covariance matrices in short windows, and the diffusion kernel reveals the dynamics by integrating those distances over the entire reference set. In this section, we propose a nonparametric Bayesian framework that exploits the chronological order to improve the extended representation. The Bayesian framework enables one to track and filter real signals sequentially without predefined statistical models. For simplicity of notation, we neglect the possible scaling between the intrinsic representation in Eq. **14** and the true underlying state space, although proper alignment or scaling is often required in practice.

In the spirit of sequential Monte Carlo methods, we represent the posterior pdf of the underlying process at time t given the underlying process at $t-1$ and a new measurement at t by a set of samples. Let $\{\boldsymbol{\theta}_t^{(k)}\}_{k=1}^{P}$ be a set of P support samples ("particles"), and let $\{w_t^{(k)}\}_{k=1}^{P}$ be a set of weights associated with the particles. The posterior pdf at t is approximated by

$$p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1},\mathbf{z}_t)\approx\sum_{k=1}^{P}w_t^{(k)}\delta(\boldsymbol{\theta}_t-\boldsymbol{\theta}_t^{(k)}), \qquad [15]$$

where the weights are denoted by

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

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

$$w_t^{(k)}\propto p\left(\boldsymbol{\theta}_t^{(k)}\middle|\boldsymbol{\theta}_{t-1}\right)p\left(\mathbf{z}_t\middle|\boldsymbol{\theta}_t^{(k)}\right). \qquad [16]$$

The densities in Eq. **16** are estimated based on the embedded domain by assuming that the local distributions in the embedded domain are good approximations to the local distributions in the true underlying process domain. Quantities associated with the embedding are denoted with a tilde (e.g., $\tilde{\boldsymbol{\theta}}_t$ denotes the embedding associated with $\boldsymbol{\theta}_t$).

We use the extended representation to approximate the likelihood function locally as a normal distribution

$$p(\mathbf{z}_t|\boldsymbol{\theta}_t)\propto\exp\left\{-\left(\boldsymbol{\Psi}(\mathbf{z}_t)-\tilde{\boldsymbol{\theta}}_t\right)^T\tilde{\mathbf{C}}_t^{-1}\left(\boldsymbol{\Psi}(\mathbf{z}_t)-\tilde{\boldsymbol{\theta}}_t\right)\right\}, \qquad [17]$$

where $\tilde{\mathbf{C}}_t$ is the local covariance matrix (now in the embedded domain) near $\tilde{\boldsymbol{\theta}}_t$. The likelihood represents the inaccuracy of the extension based on the geometric information and implies a local multivariate Gaussian distribution in the embedded domain.

We proceed by incorporating the empirical dynamics of past observations. 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

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

The samples in this neighborhood represent similar past states and can be used for dynamics estimation because their succeeding samples are available. We collect the set of succeeding samples (i.e., $\tilde{\boldsymbol{\theta}}_{\tau+1}$) for each $\tau\in\tilde{\mathcal{N}}_{t-1}$ and compute its mean and covariance, denoted by $\tilde{\boldsymbol{\mu}}_{t-1}^f$ and $\tilde{\mathbf{C}}_{t-1}^f$, respectively. The pdf of the dynamics of the underlying process measures the probability that the k-th particle is on the trajectory that passes through $\tilde{\boldsymbol{\theta}}_{t-1}$ at $t-1$, and is estimated by

$$p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})\propto\exp\left\{-\left(\tilde{\boldsymbol{\theta}}_t-\tilde{\boldsymbol{\mu}}_{t-1}^f\right)^T\left[\tilde{\mathbf{C}}_{t-1}^f\right]^{-1}\left(\tilde{\boldsymbol{\theta}}_t-\tilde{\boldsymbol{\mu}}_{t-1}^f\right)\right\}. \qquad [18]$$

Substituting Eqs. **16**–**18** into Eq. **15** yields an estimate of the posterior pdf at time t. Based on the estimated posterior pdf, a new set of particles can be drawn:

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

Because it is usually difficult to draw samples from the posterior pdf, we may use the assumption of a multivariate Gaussian model and draw samples from normal distributions. Alternatively, when a sufficient number of measurements are available, the particles can be nearest neighbors in the embedded domain, where the posterior pdf is used as a distance measure.

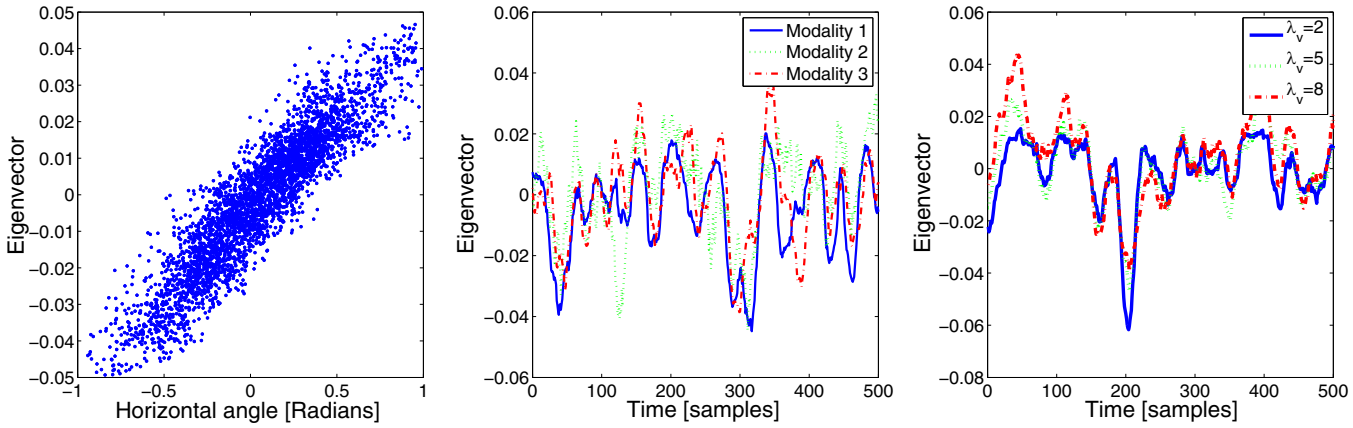

Fig. 2. Obtained embedding of the toy example. (*Left*) Scatter plot of the principal eigenvector obtained by the EIG method and the horizontal angle. (*Center*) Comparison between the obtained eigenvectors (corresponding to the vertical angle) under the three different measurement modalities (measuring the same movement). (*Right*) Comparison between the obtained eigenvectors (corresponding to the vertical angle) under modality 1 with different noise levels ($\lambda_v=2,5,8$).

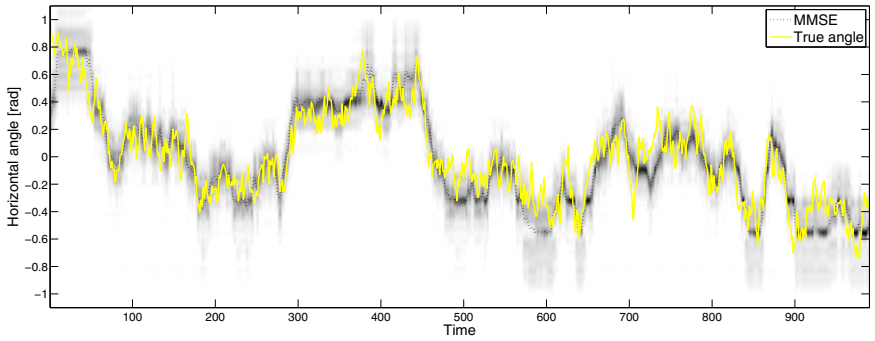

Fig. 3. Tracking the horizontal angle of the movement in a time interval following the reference segment. The yellow line is the trajectory of the horizontal angle. The vertical strips of gray-level intensity represent the posterior pdf estimate obtained by the Bayesian tracking in each time sample. The dotted black line is the expected value based on the posterior pdf estimate (MMSE estimator).

Using the estimate of the posterior pdf, an estimator of the underlying process at t can be computed according to an optimization criterion. For example, the MMSE estimator is given by

$$\hat{\boldsymbol{\theta}}_t = \mathbb{E}[\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \mathbf{z}_t] = \int \boldsymbol{\theta}_t\, p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \mathbf{z}_t) d\boldsymbol{\theta}_t \approx \sum_{k=1}^{P} p\left(\boldsymbol{\theta}_t^{(k)}\middle|\boldsymbol{\theta}_{t-1}, \mathbf{z}_t\right)\boldsymbol{\theta}_t^{(k)} = \sum_{k=1}^{P} w_t^{(k)}\boldsymbol{\theta}_t^{(k)}. \quad [19]$$

In case the particles are nearest neighbors, their estimates are already computed in previous time steps. However, for initial estimates in Eq. **19**, we require a few samples of the underlying process to align the embedding of each particle $\tilde{\boldsymbol{\theta}}_t^{(k)}$ with the true value $\boldsymbol{\theta}_t^{(k)}$.

We note that the likelihood and empirical dynamics functions based on the embedded domain provide a foundation for prediction, 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.

Application of EIG to the Toy Example

We simulated 2D trajectories of the two independent underlying angles according to Eq. **3** and the corresponding noisy measurements under the three measurement modalities. A short segment of a 2D trajectory and the corresponding 3D trajectory on the sphere are depicted in Fig. 1.

An initial interval of $N = 2{,}000$ measurements $\overline{\mathbf{z}}_s,\ s = 1, \ldots, N$ is used as the reference set, which was shown empirically to be a sufficient amount of data to represent the low-dimensional model of the angles. We note that of the 2,000 reference measurements, merely the first 100 had known angles, which were used to align the embedded samples $\tilde{\boldsymbol{\theta}}_t$ to the true samples $\boldsymbol{\theta}_t$. For each reference measurement, we computed a histogram with $m = 20$ bins in a short window of $N_h = 20$ samples, which approximated the pdf, and the corresponding local covariance matrix according to Eq. **6** for $N_c = 5$. Next, the kernel was calculated by Eq. **9**. Finally, the embedding of the reference measurements was constructed based on the spectral decomposition of the kernel by Eq. **10**.

Measurements $\mathbf{z}_t$ at times $t > 2{,}000$ were sequentially processed. For each measurement, the kernel based on Eq. **11** and the embedding in Eq. **14** were computed. The posterior pdf was then estimated according to the proposed Bayesian framework by relying on the embedded domain. Finally, the MMSE estimator of the underlying process at each time step t was calculated according to Eq. **19**.

Fig. 2 (*Left*) shows a scatter plot of the obtained embedding of measurements under modality 1 using the EIG method. The coordinates of the principal eigenvector are presented as a function of the samples of the horizontal angle. We observe that the principal eigenvector obtained via the proposed method is linearly correlated with the horizontal angle. We note that a similar correlation was found between the eigenvector associated with the second largest eigenvalue and the vertical angle. A MATLAB (MathWorks) code that implements this toy example is available online (http://www.runmycode.org/).

The same object movement was measured via the three modalities, and the embedding of each set of measurements was computed. Fig. 2 (*Center*) depicts the three eigenvectors (each associated with the second largest eigenvalue and obtained under a different measurement modality) as a function of time t. We note that the presented 500 coordinates of the eigenvectors were computed by extension and represent the underlying angle of $\mathbf{z}_t$, $t > 2{,}000$. The measurements under the three modalities are very different in their nature (e.g., spike sequences in modality 1 and modality 2 compared with noisy 3D coordinates in modality 3). We observe eigenvectors with similar trends, which implies intrinsic modeling and demonstrates the invariance of the proposed approach to the type of the measurement modality and noise. To demonstrate the resilience to measurement noise further, we repeated the experiment and measured the same movement via modality 1 under various noise conditions. In Fig. 2 (*Right*), we present the obtained three eigenvectors and observe similar trends.

Fig. 3 presents the Bayesian tracking of the vertical angle under modality 1 in a time interval of 1,000 samples following the reference segment. The yellow line (Fig. 3) is the true trajectory of the vertical angle. The vertical strips of gray level intensity (Fig. 3) represent the posterior pdf estimate obtained by the Bayesian tracking in each time sample. The dotted black line (Fig. 3) is the expected value based on the posterior pdf estimate. The approach based on EIG was compared with the Bootstrap particle filter

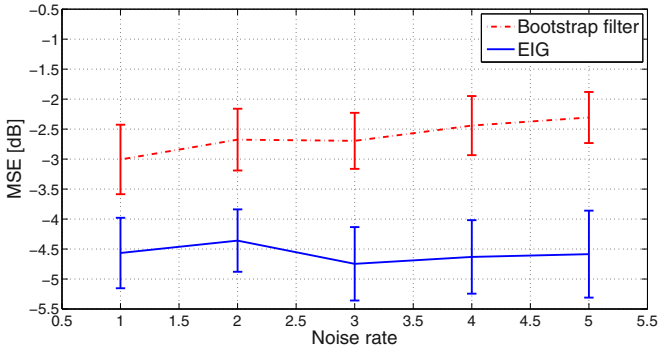

Fig. 4. The mean square error (MSE) obtained under modality 1 as a function of the noise level λ_v.

(22), with resampling 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 as well as the noise statistics as priors, whereas the EIG approach uses past observations. To compare the two algorithms, we used the same number of particles ($P = 50$). The results under modality 1 with different noise levels are summarized in Fig. 4. The results were averaged over several realizations of measurements and noise as well as over several repetitions of the stochastic particle filter tracking. We note that the typical spike rate ranged between 5 and 15 and that the measurement noise rate ranged between 0 and 5. Thus, we used 20 histogram bins equally divided in $[0, 20]$. We observe that the EIG 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 Bootstrap particle filter uses an importance sampling measure that ignores the new measurement. In this case, the particle filter, which relies on the true statistical model, may suffer from overfitting of the model to the available *noisy* data and is shown to be inferior to the data-driven EIG method. In addition, the performance of the particle filter degrades as the noise level increases, whereas the performance of the EIG method is less sensitive to changes in the noise level.

Acoustic Localization

We revisit the experiment presented by Talmon et al. (23, 24) and test the ability of the EIG method to recover the location of a sound source (25) from a *single microphone* in a room. Following is a brief description of the experiment setup. Inside a reverberant room, we placed 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 different realization of a zero-mean and unit variance white Gaussian noise was played from the speaker and recorded into 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, because the rest of the room parameters are fixed. Mild variations caused by the movement of the arm are neglected. In this work, we calculated histograms of the recorded raw signal in short-time windows and computed their corresponding local covariance matrices. We then built the kernel and obtained the embedding. Similar to the results reported by Talmon et al. (24), the leading eigenvector represents the horizontal angle. The recovery of the angle based on the embedding attained an average error of 2°. This result is slightly inferior to the result obtained in by Talmon et al. (24). However, in the study by Talmon et al. (24), a priori knowledge of noiseless Gaussian recordings was used. In addition, the features were computed at several scales. In this work, no prior knowledge needed to be supplied and the short-term histograms were computed from the raw data.

Summary

In this paper, we have presented a method for intrinsic modeling of time series without a priori knowledge. This method extends existing manifold learning techniques to support the stochastic settings of nonlinear filtering problems. We showed that the inferred model is intrinsic, that is, invariant to the measurement modality and noise. Based on this model, we developed a nonparametric Bayesian framework that enables one to filter signals optimally without existing definitive models, and hence may be suitable for real-world applications. In future work, we intend to develop error estimates and criteria for signal rates as well as to explore the choice of proper configuration (e.g., the choice of window lengths). In addition, we plan to apply the presented approach to biomedical and neuroscience applications, such as the decoding of neural data in the human auditory and visual systems.

ACKNOWLEDGMENTS. R.T. is partially funded by the Viterbi Fellowship Technion - Israel Institute of Technology.

1. Tenenbaum JB, de Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. *Science* 290(5500):2319–2323.
2. Donoho DL, Grimes C (2003) Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proc Natl Acad Sci USA* 100(10):5591–5596.
3. Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput* 15(6):1373–1396.
4. Coifman RR, Lafon S (2006) Diffusion maps. *Appl Comput Harmon Anal* 21:5–30.
5. Singer A, Coifman RR (2008) Non-linear independent component analysis with diffusion maps. *Appl Comput Harmon Anal* 25(2):226–239.
6. Laffery JD, Lebanon G (2005) Diffusion kernels on statistical manifolds. *J Mach Learn Res* 6:129–163.
7. Amari S, Nagaoka H (2007) *Methods of Information Geometry* (American Mathematical Society, Providence, RI).
8. Doucet A, Godsill S, Andrieu C (2000) On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat Comput* 10(3):197–208.
9. Arulampalam MS, Maskell S, Gordon N, Clapp T (2002) A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans Signal Process* 50(2): 174–188.
10. Bar-Shalom Y, Fortmann TE (1988) *Tracking and Data Association* (Academic Press, Boston).
11. Singer A, Erban R, Kevrekidis IG, Coifman RR (2009) Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps. *Proc Natl Acad Sci USA* 106(38):16090–16095.
12. Wiskott L, Sejnowski TJ (2002) Slow feature analysis: unsupervised learning of invariances. *Neural Comput* 14(4):715–770.
13. Storvik G (2002) Particle filters for state-space models with the presence of unknown static parameters. *IEEE Trans Signal Process* 50(2):281–289.
14. Godsill SJ, Vermaak J, Ng W, Li JF (2007) Models and algorithms for tracking of maneuvering objects using variable rate particle filters. *Proc IEEE* 95(5):925–952.
15. Girolami M, Calderhead B (2011) Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J R Stat Soc Series B Stat Methodol* 73:123–214.
16. Talmon R, Coifman RR (2012) Empirical intrinsic modeling of signals and information geometry. Technical Report YALEU/DCS/TR-1467. Available at http://www.cs.yale.edu/publications/techreports/tr1467.pdf. Accessed November 9, 2012.
17. Hein M, Audibert Y (2005) Intrinsic dimensionality estimation of submanifolds in $\mathbb{R}^d$. *Proc Int Conf Mach Learn* 289–296.
18. Chung FRK (1997) *Spectral Graph Theory. CBMS Regional Conference Series in Mathematics* (American Mathematical Society, Providence, RI).
19. Kushnir D, Haddad A, Coifman RR (2011) Anisotropic diffusion on sub-manifolds with application to earth structure classification. *Appl Comput Harmon Anal* 32(2): 280–294.
20. Singer A (2006) Spectral independent component analysis. *Appl Comput Harmon Anal* 21(1):135–144.
21. Talmon R, Cohen I, Gannot S, Coifman RR (2012) Supervised graph-based processing for sequential transient interference suppression. *IEEE Trans Audio Speech Lang Processing* 20(9):2528–2538.
22. Gordon N, Salmond D, Smith AFM (1993) Novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE Proc F* 140(2):107–113.
23. Talmon R, Kushnir D, Coifman RR, Cohen I, Gannot S (2012) Parametrization of linear systems using diffusion kernels. *IEEE Trans Signal Process* 60(3):1159–1173.
24. Talmon R, Cohen I, Gannot S (2011) Supervised source localization using diffusion kernels. *Proceedings of the IEEE Workshop on Application of Signal Processing to Audio and Acoustics* 245–248.
25. Huang Y, Benesty J, Chen J (2006) *Acoustic MIMO Signal Processing* (Springer, Berlin).