Nonlinear Modeling and Processing Using Empirical Intrinsic Geometry with Application to Biomedical Imaging

Ronen Talmon[1], Yoel Shkolnisky[2], and Ronald R. Coifman[1]

[1] Mathematics Department, Yale University
[2] Applied Mathematics Department, Tel Aviv University
ronen.talmon@yale.edu

Abstract. In this paper we present a method for intrinsic modeling of nonlinear filtering problems without a-priori knowledge using empirical information geometry and empirical differential geometry. We show that the inferred model is noise resilient and invariant under different random observations and instrumental modalities. In addition, we show that it can be extended efficiently to newly acquired measurements. Based on this model, we present a Bayesian framework for nonlinear filtering, which enables to optimally process real signals without predefined statistical models. An application to biomedical imaging, in which the acquisition instruments are based on photon counters, is demonstrated; we propose to incorporate the temporal information, which is commonly ignored in existing methods, for image enhancement.

Keywords: Intrinsic model, differential geometry, information geometry, nonparametric estimation, nonlinear dynamical systems.

1 Introduction

Nonlinear filtering problems are usually formulated in the state-space, which naturally introduces an underlying process on a manifold. The state-space formalism includes two models: a (possibly nonlinear) dynamical model which consists of a stochastic differential equation describing the evolution of the underlying process (state) with time, and the (nonlinear) measurement model that relates the noisy observations to the underlying process. The prior knowledge of the two models is essential in estimation problems. Specifically, it is required in Bayesian algorithms, e.g. the Kalman Filter and its extensions [3] as well as contemporary sequential Monte Carlo algorithms [2,6]. Unfortunately, these models might be unknown and difficult to reveal in real applications.

In this paper, our main goal is to provide viable models to such nonlinear signal processing problems. We present a method to construct *intrinsic* models, which are invariant to the measurement modality and noise. In addition, the obtained models can be extended to new measurements in a sequential manner. We adopt ideas from information geometry, which provides a convenient framework

F. Nielsen and F. Barbaresco (Eds.): GSI 2013, LNCS 8085, pp. 441–448, 2013.

to combine geometric and statistical analysis, and obtain a representation of empirical distributions of signals. However, unlike traditional information geometry, we infer the underlying model of distributions from measurements instead of using known models. The intrinsic model is obtained through eigenvectors of an appropriate Laplace operator [4,5]. The role of the Laplace operator is to quantify the connections between the measurements and to integrate all the information. Specifically, its eigenvectors provide an embedding (or a parametrization) of the measurements, which is viewed as a representation of the underlying process on the parametric manifold. Based on the inferred intrinsic models, we present a nonparametric Bayesian framework for tracking and estimation. This framework enables to optimally process real signals without statistical descriptions nor predefined models.

We believe that such a nonparametric Bayesian framework may be useful for biological imaging. An imaging process usually consists of the emission of radiation or light on an object or a specimen of interest for a certain time interval. The corresponding output signal of the acquisition sensors (e.g., photon counters) is a time series of the instantaneous amount of radiation at the sensors that travelled through the object of interest over the duration of the imaging process. Since such an imaging process is known to be very noisy, common practice is to compute the mean value of the time series at each sensor in order to suppress the (zero-mean) additive noise. In addition, in order to obtain better suppression, the duration, and hence, the amount of radiation, is extended as much as possible. In case the sensors are positioned in a 2-d grid array, the means of the outputs of the sensors yield a 2-d image, in which each pixel corresponds to a sensor. Unfortunately, the object may move or vibrate (or even change due to exposure to radiation) during the acquisition process, and each sensor might capture descriptions of different parts of the object. Thus, computing the mean of each time series should include a proper alignment according to the movement of the object. We propose to exploit the temporal information (which is usually discarded during simple averaging) and to use the presented nonparametric Bayesian framework to empirically reveal and track the movement of the object. Tracking the movement enables to align the time series across the sensors, and hence, to obtain better noise suppression. This may provide a better image quality and may help to reduce the exposure time to radiation. In this paper we demonstrate this approach on a simple simulated imaging model.

2 Problem Formulation

Let $\boldsymbol{\theta}_t$ be a d-dimensional underlying process in time index t. The dynamics of the process are described by normalized stochastic differential equations as follows[1]

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

[1] x^i denotes the i-th coordinate of a vector $\mathbf{x}$.

where a^i are unknown drift functions and $\dot{w}_t^i$ are independent white noises. We note that the underlying process is equivalent to the system state in the classical terminology.

Let $\mathbf{z}_t$ denote an n-dimensional measurement process in time t, given by

$$\mathbf{z}_t = g(\mathbf{y}_t, \mathbf{v}_t), \tag{2}$$

where g is an unknown (possibly nonlinear) measurement function, $\mathbf{y}_t$ is a "clean" measurement obtained in noiseless conditions, and $\mathbf{v}_t$ is a corrupting n-dimensional measurement noise. We assume that $\mathbf{y}_t$ is drawn from a probability density function (pdf) $f(\mathbf{y}; \boldsymbol{\theta})$, thereby the statistics of the measurement process at time t are time-varying and depend on the underlying process $\boldsymbol{\theta}$ at time t. In addition, $\mathbf{v}_t$ is drawn from an unknown stationary pdf $q(\mathbf{v})$ and is independent of $\mathbf{y}_t$.

The state $\boldsymbol{\theta}_t$ constitutes a parametric manifold that is transformed into the observable manifold of measurements. Our goal in this work is to recover $\boldsymbol{\theta}_t$ and its dynamics based on a sequence of measurements $\{\mathbf{z}_t\}$.

3 Local Probability Models

The time-varying pdf of the measured process $\mathbf{z}_t$ is a function of $\boldsymbol{\theta}_t$, and hence, consists of important information. Unfortunately, the pdfs are unknown and we can only use the empirical distributions. Let $\mathbf{h}_t$ be an m-bins histogram of the measurements in a short window centered at time t, whose j-th element is approximated by

$$h_t^j = \int_{\mathbf{z} \in \mathcal{H}_j} p(\mathbf{z}; \boldsymbol{\theta}) d\mathbf{z}, \tag{3}$$

where $p(\mathbf{z}; \boldsymbol{\theta})$ denotes the pdf of the measured process $\mathbf{z}_t$ and $\mathcal{H}_j$ is the j-th bin.

Lemma 1. *The pdf $p(\mathbf{z}; \boldsymbol{\theta})$ of the measured process $\mathbf{z}_t$ is given by a linear transformation of the pdf $f(\mathbf{y}; \boldsymbol{\theta})$ of the clean measurement component $\mathbf{y}_t$.*

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

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

Combining (3) and Lemma 1, 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. We view the histograms as features of the measurements.

From (1) and by using Itô lemma, it was shown in [8] that the (j,k)-th element of the $m \times m$ covariance matrix $\mathbf{C}_t$ of $\mathbf{h}_t$ is given by

$$C_t^{jk} = \mathrm{Cov}(h_t^j, h_t^k) = \sum_{i=1}^{d} \frac{\partial h^j}{\partial \theta^i} \frac{\partial h^k}{\partial \theta^i} = \sum_{i=1}^{d} J_t^{ji} J_t^{ki}, \quad j,k = 1,\dots,m, \tag{5}$$

where $\mathbf{J}_t$ is the $m \times d$ Jacobian matrix of $\mathbf{h}_t$. In matrix form, (5) can be rewritten as $\mathbf{C}_t = \mathbf{J}_t \mathbf{J}_t^T$, which yields that the covariance matrix $\mathbf{C}_t$ is a semi-definite positive matrix of rank d.

We use the histograms and their corresponding covariance matrices to describe the local statistical model $\mathcal{Z}_t$ for each measurement, which is assumed to be a Gaussian centered at $\mathbf{h}_t$ with $\mathbf{C}_t$ covariance, i.e., $\mathcal{N}(\mathbf{h}_t, \mathbf{C}_t)$. In practice, the empirical covariance matrices are computed in short time windows around each histogram $\mathbf{h}_t$. We note that these local models serve as an intermediate step in inferring the intrinsic model of the problem.

4 Intrinsic Modeling

Let $\{\bar{\mathbf{z}}_s\}_{s=1}^N$ be a sequence of N reference measurements. For this sequence we estimate the local densities and their covariance matrices as well as the local Gaussian models $\{\mathcal{Z}_s\}$. This enables us to define a non symmetric kernel $\mathbf{A}$ between any measurement $\mathbf{z}_t$ and the N reference measurements as

$$A^{ts} = \Pr(\mathbf{z}_t | \mathbf{z}_t \in \mathcal{Z}_s) = (2\pi)^{m/2} |\mathbf{C}_s|^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{h}_t - \mathbf{h}_s)^T \mathbf{C}_s^{-1} (\mathbf{h}_t - \mathbf{h}_s) \right\}. \tag{6}$$

Consider the kernel $\mathbf{W}_r = \mathbf{A}^T \mathbf{A}$ of the reference measurements. It was shown in [7] that up to a normalization its (s,s')-th element is the following affinity measure between $\bar{\mathbf{z}}_s$ and $\bar{\mathbf{z}}_{s'}$ given by

$$W_r^{ss'} = \exp\left\{ -d^2(\bar{\mathbf{z}}, \bar{\mathbf{z}}_{s'}) \right\}, \tag{7}$$

and

$$d^2(\bar{\mathbf{z}}_s, \bar{\mathbf{z}}_{s'}) = (\mathbf{h}_s - \mathbf{h}_{s'})^T \left(\mathbf{C}_s + \mathbf{C}_{s'} \right)^{-1} (\mathbf{h}_s - \mathbf{h}_{s'}) \tag{8}$$

is the *Mahalanobis* distance. Since the Mahalanobis distance is invariant under linear transformations, by Corollary 1 it is *invariant to any measurement noise*. By assuming $\mathbf{h}_s = h(\boldsymbol{\theta}_s)$ is a bi-Lipschitz smooth function of the underlying process $\boldsymbol{\theta}_s$ and by using local linearization of the function, i.e., $\mathbf{h}_s = \mathbf{J}_s^T \boldsymbol{\theta}_s + \epsilon_s$, it was shown by Singer and Coifman in [8] that the Mahalanobis distance approximates the Euclidean distance between the corresponding samples of the underlying process to a second order, i.e.,

$$\|\boldsymbol{\theta}_s - \boldsymbol{\theta}_{s'}\|^2 = d^2(\bar{\mathbf{z}}_s, \bar{\mathbf{z}}_{s'}) + O(\|\mathbf{h}_s - \mathbf{h}_{s'}\|^4). \tag{9}$$

This result implies that the Mahalanobis distance is *invariant to the measurement modality.*

Consider now the "dual" kernel $\mathbf{W} = \mathbf{A}\mathbf{A}^T$. It can be shown that its (t, t')-th element consists of an affinity measure which is equal to the probability that any two measurements are associated with the same local probability model [9], i.e.,

$$W^{tt'} = \Pr(\mathbf{z}_t \in \mathcal{Z}_s, \mathbf{z}_{t'} \in \mathcal{Z}_s | \mathbf{z}_t, \mathbf{z}_{t'}). \tag{10}$$

By [7,8], $\mathbf{W}$ converges to a diffusion operator when we have sufficient amount of measurements and the local models are defined in infinitesimal neighborhoods. Such an operator reveals the low-dimensional underlying manifold, for which the eigenvectors give an approximate parametrization. Thus, we compute the eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\boldsymbol{\psi}_i\}$ of $\mathbf{W}$. In the case of a flat manifold, we assume that the leading d eigenvectors recover d proxies for the underlying process up to a monotonic scaling [7]. We define a d-dimensional representation by

$$\boldsymbol{\Psi}(\mathbf{z}_t) \triangleq \left[\psi_1^t, \psi_2^t, \ldots, \psi_d^t\right], \tag{11}$$

for each measurement at time t. Then, the embedding (11) is seen as the obtained modeling of the measurements revealing the corresponding underlying process.

The aforementioned construction of the embedding is especially suitable for sequential extension [9] consisting of two stages: a training stage in which a sequence of training measurements is assumed to be available in advance, and a test stage in which new incoming measurements are sequentially processed.

In the training stage, reference measurements are processed to form a learned model. The feature vectors (histograms) and the corresponding local covariance matrices are computed. The kernel $\mathbf{W}_r$ is directly computed and its eigen-decomposition is calculated. The eigenvectors of the kernel form a learned model for the training set.

In the test stage, as new incoming measurements become available, we construct $\mathbf{A}$ according (6), and then, compute the extended representation by exploiting the relationship between the kernels $\mathbf{W}_r$ and $\mathbf{W}$, given by the singular value decomposition of $\mathbf{A}$ as

$$\boldsymbol{\psi}_i = \frac{1}{\sqrt{\lambda_i}} \mathbf{A} \boldsymbol{\varphi}_i, \tag{12}$$

where $\boldsymbol{\varphi}_j$ are the eigenvectors of $\mathbf{W}_r$. It is worthwhile noting that the extension does not involve an eigen-decomposition but rather relies on the an algebraic relationship. Thus, the processing of new measurements involves low computational complexity [9].

Relationship to Information Geometry. In classical information geometry [1], the parameters of the distribution of the measurements confine the data to an underlying manifold. The distribution is usually required in an analytic form and the Kullback Liebler (KL) divergence is used as a comparison metric. In this paper, we present a data-driven approach to recover the manifold without a priori knowledge. Instead, we propose to rely on empirical distributions and use the Mahalanobis distance as a metric between distributions to obtain their intrinsic parameterization. The Mahalanobis distance has a tight relationship to the KL divergence and uses a local Gaussian distribution assumption. We emphasize that the local Gaussian assumption is merely an intermediate step in obtaining the intrinsic global parameterization. For more details see [11].

5 Bayesian Tracking

In this section we incorporate the dynamics of time series, which was neglected thus far in the intrinsic model computation. We present a nonparametric Bayesian framework [10] that enables to filter real signals without predefined statistical models. For simplicity of notation, we neglect the possible scaling between the intrinsic representation in (11) and the true underlying state space, although in practice, proper alignment or scaling is often required.

We use (11) to locally approximate the likelihood function as the following normal distribution

$$\Pr(\mathbf{z}_t|\boldsymbol{\theta}_t) \propto \exp\left\{-\left(\boldsymbol{\Psi}(\mathbf{z}_t)-\boldsymbol{\theta}_t\right)^T \mathbf{C}_{\boldsymbol{\theta},t}^{-1}\left(\boldsymbol{\Psi}(\mathbf{z}_t)-\boldsymbol{\theta}_t\right)\right\}, \tag{13}$$

where $\boldsymbol{\theta}_t$ is assumed to be the true underlying sample and $\mathbf{C}_{\boldsymbol{\theta},t}$ is the local covariance (now in the state-space) near $\boldsymbol{\theta}_t$.

We proceed by incorporating the empirical dynamics of past observations as a prior. Let $\mathcal{N}_{t-1}$ be a set of time indices of samples in a $\xi > 0$ neighborhood of $\boldsymbol{\theta}_{t-1}$, defined as

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

The samples in this neighborhood represent similar past states and can be used for dynamics estimation since their succeeding samples are available. We collect the succeeding samples, i.e., $\boldsymbol{\theta}_{s+1}$ for each $s \in \mathcal{N}_{t-1}$, and compute their mean and covariance, denoted by $\bar{\boldsymbol{\theta}}_{\boldsymbol{\theta},t-1}^f$ and $\mathbf{C}_{\boldsymbol{\theta},t-1}^f$, respectively. The pdf of the dynamics of the underlying process is estimated by

$$\Pr(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) \propto \exp\left\{-\left(\boldsymbol{\theta}_t-\bar{\boldsymbol{\theta}}_{\boldsymbol{\theta},t-1}^f\right)^T\left[\mathbf{C}_{\boldsymbol{\theta},t-1}^f\right]^{-1}\left(\boldsymbol{\theta}_t-\bar{\boldsymbol{\theta}}_{\boldsymbol{\theta},t-1}^f\right)\right\}. \tag{14}$$

Since we merely have pointwise definitions of the statistical models, we use the concept of sequential Monte Carlo methods [2,6] and represent the posterior pdf by a set of support samples $\{\boldsymbol{\theta}_t^{(k)}\}_{k=1}^P$ ("particles"), i.e.,

$$\Pr(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1},\mathbf{z}_t) \approx \sum_{k=1}^{P} w_t^{(k)} \delta\left(\boldsymbol{\theta}_t - \boldsymbol{\theta}_t^{(k)}\right), \tag{15}$$

where the weights are given by

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

with $\sum_{k=1}^{P} w_t^{(k)} = 1$. We therefore have a discrete approximation of the desired posterior pdf. At each time step, the particles are drawn from the posterior pdf estimate of the preceding step. By Bayes' theorem and by the Markov dynamical model, we obtain

$$w_t^{(k)} \propto \Pr(\boldsymbol{\theta}_t^{(k)}|\boldsymbol{\theta}_{t-1}) \Pr(\mathbf{z}_t|\boldsymbol{\theta}_t^{(k)}). \tag{16}$$

The densities in (16) are estimated based on the embedded domain from (13) and (14). Using the estimate of the posterior pdf, a sequential estimator of the underlying process at t can be computed according to an optimization criterion. For example, the minimum mean squared error (MMSE) estimator is given by

$$\hat{\boldsymbol{\theta}}_t = \mathbb{E}\left[\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \mathbf{z}_t\right] = \int \boldsymbol{\theta}_t p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \mathbf{z}_t) d\boldsymbol{\theta}_t \approx \sum_{k=1}^{P} w_t^{(k)} \boldsymbol{\theta}_t^{(k)}. \tag{17}$$

We emphasize that the Gaussian prior and likelihood represent the distribution in the low dimensional inferred space and are used merely for tracking.

6 Application to Imaging

In the following experimental study, we simulate a simple imaging model of a 2-d shape measured by a 1-d linear sensor array. Assume we measure the shape of a rigid biological material that vibrates over time. Let θ_t denote the position of the center of the object, which is assumed to be a diffusion process with unknown coefficients (1). For simplicity, we set the initial position at the origin. The noisy measurements $\mathbf{z}_t = \mathbf{y}_t + \mathbf{v}_t$ are the output signals of 15 simulated sensors, where $\mathbf{v}_t$ is a corrupting white Gaussian noise. In this case, the clean measurement component in the i-th sensor is given by

$$y_t^i = f(p_i - \theta_t)$$

where p_i is the location of the i-th sensor on the array axis, and f is the distance that the beam of radiation or light traveled through the object. Figure 1(a) depicts the experimental setup. The objective in this experiment is to suppress the noise in the measurements $\mathbf{z}_t$ and obtain an estimate of the measured shape $f(p_i)$ at each sensor at $t = 0$. In this simulation, we chose f to be a Gaussian.

We applied the presented Bayesian framework to reveal and track θ_t from the noisy measurements without any additional information on the models, which were used merely to simulate the data. Figure 1(b) presents the tracking results using the MMSE estimator (17). This result exemplifies the tracking ability of the presented Bayesian framework, which solely relies on the measured signal.

Next, we utilize the recovered movement for aligned averaging to improve the noise suppression. Let $\hat{f}_u(p_i) = \sum_t z_t^i$ and $\hat{f}_a(p_i) = \sum_t z_t^{\hat{i}_t}$ be estimates of the shape obtained by unaligned and aligned averaging, respectively, where $\hat{i}_t = \arg\min_j |p_j - (p_i + \hat{\theta}_t)|$ and $\hat{\theta}_t$ is the MMSE estimate of the position of the center. We computed the mean square error (MSE) between these estimates and the true measured shape $f(p_i)$ as an objective performance measure. The aligned

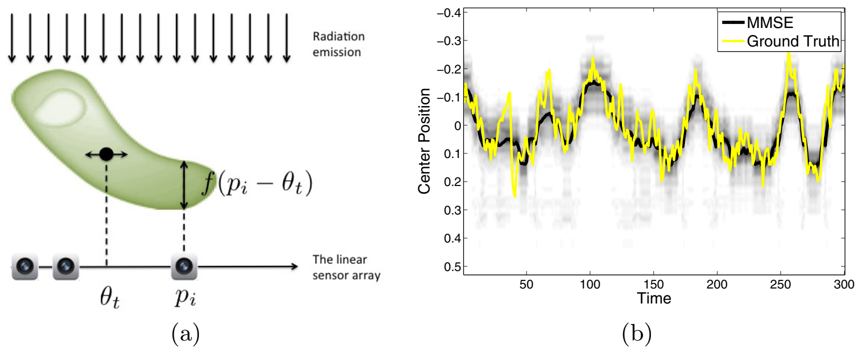

Fig. 1. (a) An illustration of the experimental setup. (b) Tracking the center position of the shape. The yellow line is the true position of the center. The vertical strips of gray level intensity represent the posterior pdf estimate obtained by the Bayesian tracking in each time sample. The solid black line is the expected value based on the posterior pdf estimate (MMSE estimator (17)).

estimator showed 11.2 dB mean improvement over the MSE obtained by the unaligned estimator. This result implies on the potential benefit of incorporating temporal information and nonparametric Bayesian tracking in imaging.

References

1. Amari, S., Nagaoka, H.: Methods of information geometry. American Mathematical Society 191 (2007)
2. Arulampalam, M.S., Maskell, S., Gordon, N., Clapp, T.: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. Signal Process. 50, 174–188 (2003)
3. Bar-Shalom, Y., Fortmann, T.E.: Tracking and data association. Academic Press (1988)
4. Belkin, M., Niyogi, P.: Laplacian Eigenmaps for dimensionality reduction and data representation. Neural Computation 15, 1373–1396 (2003)
5. Coifman, R.R., Lafon, S.: Diffusion maps. Appl. Comp. Harm. Anal. 21, 5–30 (2006)
6. Doucet, A., Godsill, S., Andrieu, C.: On sequential Monte Carlo sampling methods for Bayesian filtering. Statistics and Computing 10, 197–208 (2000)
7. Kushnir, D., Haddad, A., Coifman, R.R.: Anisotropic diffusion on sub-manifolds with application to earth structure classification. Appl. Comp. Harm. Anal. 32, 280–294 (2012)
8. Singer, A., Coifman, R.R.: Nonlinear independent component analysis with diffusion maps. Appl. Comp. Harm. Anal. 25, 226–239 (2008)
9. Talmon, R., Cohen, I., Gannot, S., Coifman, R.R.: Supervised graph-based processing for sequential transient interference suppression. IEEE Trans. Audio Speech Lang. Process. 20(9), 2528–2538 (2012)
10. Talmon, R., Coifman, R.R.: Empirical Intrinsic Geometry for Nonlinear Modeling and Time Series Filtering. Yale-CS TR, 1451 (2012)
11. Talmon, R., Coifman, R.R.: Empirical intrinsic modeling of signals and information geometry. Yale-CS TR, 1467 (2012)