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

Ronen Talmon\textsuperscript{1}, Yoel Shkolnisky\textsuperscript{2}, and Ronald Coifman\textsuperscript{1}

\textsuperscript{1}Mathematics Department, Yale University
\textsuperscript{2}Applied Mathematics Department, Tel Aviv University

Geometric Science of Information (GSI 2013)
August 28-30, 2013, Paris
• Example for Intrinsic Modeling I

- **Molecular Dynamics**
- Consider a molecule oscillating stochastically in water
  - For example, Alanine Dipeptide
- Due to the coherent structure of molecular motion, we assume that the configuration at any given time is essentially described by a small number of structural variables
  - In the Alanine case, we will discover two factors, corresponding to the dihedral angles
• Example for Intrinsic Modeling I

- We observe three atoms of the molecule for a certain period, three other atoms for a second period, and the rest in the last period.

- **The task** is to describe the positions of all atoms at all times:
  - More precisely, derive intrinsic variables that correspond to the dihedral angles and describe their relation to the positions of all atoms.
  - We always derive the same intrinsic variables (angles) from partial observations (independently of the specific atoms we observe).
  - If we learn the model, we can describe the positions of all atoms.
• Example for Intrinsic Modeling II

- Predicting Epileptic Seizures
- **Goal**: to warn the patient prior to the seizure (when medication or surgery are not viable)
- **Data**: intracranial EEG recordings
• **Example for Intrinsic Modeling II**

- **Our assumption**: the measurements are controlled by underlying processes that represent the brain activity

- **Main Idea**: predict seizures based on the “brain activity processes”

- **Challenges**: Noisy data, unknown model, and no available examples
• **Manifold Learning**

- Represent the data as points in a high dimensional space
- The points lie on a low dimensional structure (manifold) that is governed by latent factors
- For example, atom trajectories and the dihedral angles
Introduction

• Manifold Learning

- Traditional manifold learning techniques:
  - Laplacian eigenmaps [Belkin & Niyogi, 03’]
  - Diffusion maps [Coifman & Lafon, 05’; Singer & Coifman, 08’]
Dynamical model: let $\theta_t$ be a $d$-dimensional underlying process (the state) in time index $t$ that evolves according to

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

where $a^i$ are unknown drift coefficients and $w_t^i$ are independent white noises.

Measurement modality: let $z_t$ be an $n$-dimensional measured signal, given by

$$z_t = g(y_t, v_t)$$

- $y_t$ is the clean observation component drawn from the time-varying pdf $f(y; \theta)$
- $v_t$ is a corrupting noise (independent of $y_t$)
- $g$ is an arbitrary measurement function

The goal: recover and track $\theta_t$ given $z_t$
Manifold Learning for Time Series

- The general outline:
  - Construct an affinity matrix (kernel) between the measurements $z_t$, e.g.,
    \[ k(z_t, z_s) = \exp\left\{-\frac{\|z_t - z_s\|^2}{\varepsilon}\right\} \]
  - Normalize the kernel to obtain a Laplace operator [Chung, 97']
  - The spectral decomposition (eigenvectors) represents the underlying factors
The mapping between the observable data and the underlying processes is often stochastic and contains measurement noise

- Repeated observations of the same phenomenon usually yield different measurement realizations
- The measurements may be performed using different instruments/sensors

Each set of related measurements of the same phenomenon will have a different geometric structure

- Depending on the instrument and the specific realization
- Poses a problem for standard manifold learning methods
Intrinsic Modeling

![Diagram of Intrinsic Modeling](image)
• How to Obtain an Intrinsic Model?

Q: Does the Euclidean distance between the measurements convey the information?
Realizations of a random process and measurement noise

\[ k(z_t, z_s) = \exp \left( \frac{||z_t - z_s||^2}{\varepsilon} \right) \]

A: We propose a new paradigm - **Empirical Intrinsic Geometry (EIG)**

[Talmon & Coifman, PNAS, 13’]
- Find a proper high dimensional representation
- Find an **intrinsic** distance measure: robust to measurement noise and modality
• Geometric Interpretation

- Exploit perturbations to explore and learn the tangent plane
- Compare the points based on the principal directions of the tangent planes ("local PCA")
The Mahalanobis Distance

- We view the local histograms as feature vectors for each measurement
  \[ z_t \rightarrow h_t \]

- For each feature vector, we compute the local covariance matrix in a temporal neighborhood of length \( L \)
  \[ C_t = \frac{1}{L} \sum_{s=t-L+1}^{t} (h_s - \mu_t)(h_s - \mu_t)^T \]

  where \( \mu_t \) is the local mean

Definition – Mahalanobis Distance

- Define a symmetric \( C \)-dependent distance between feature vectors
  \[ d_C^2(z_t, z_s) = \frac{1}{2} (h_t - h_s)^T (C_t^{-1} + C_s^{-1})(h_t - h_s) \]
• Results

Assumption

- The histograms are linear transformations of the pdf $p(z; \theta)$

- Each histogram bin can be expressed as

$$h_i^j = \int_{z \in \mathcal{H}_j} p(z; \theta) \, dz$$

where $\mathcal{H}_j$ are the histogram bins

Lemma

- In the histograms domain, any stationary noise is a linear transformation

- By relying on the independence of the processes:

$$p(z; \theta) = \int f(y; \theta) q(v) \, dy dv$$

$$g(y, v) = z$$
The Mahalanobis distance:

- Is invariant under linear transformations, thus by lemma, noise resilient
- Approximates the Euclidean distance between samples of the underlying process, i.e.,

\[ \|\theta_t - \theta_s\|^2 = d_C^2(z_t, z_s) + O(\|h_t - h_s\|^4) \]

Assumption

- The process \(h_t\) can be described as a (possibly nonlinear) bi-Lipschitz function of the underlying process \(\theta_t\)
- We rely on a first order approximation of the measurement function:

\[ h_t = J_t^T \theta_t + \epsilon_t \]

where \(J_t\) is the Jacobian, defined as \(J_t^{ji} = \frac{\partial h^j}{\partial \theta^i} \)
Q: Does the structure of the measurements convey the information?

A: The local densities of the measurements do and not particular realizations

- **Information Geometry** [Amari & Nagaoka, 00’]:
  - Use the **Kullback-Liebler divergence** approximated by the Fisher metric
    \[
    D(p(z_t; \theta) || p(z_{t_0}; \theta)) = \delta \theta_t^T I_t \delta \theta_t
    \]
    where \(I_t\) is the **Fisher Information matrix**

- **EIG**: a similar data-driven metric: consider the following features
  \[
  l_t^j = \alpha_j \log (h_t^j)
  \]

**Theorem**

- \(I_t = J_t^T J_t\) (underlying manifold dimensionality)
- \(C_t = J_t J_t^T\) (feature vectors dimensionality)
• Anisotropic Kernel

- Let \( \{ \mathbf{z}_t \}_{t=1}^N \) be a set of measurements
  
  - For each measurement, we compute the local histogram and covariance

- Construct an \( N \times N \) symmetric affinity matrix

  \[
  W^{ts} = c \exp \left\{ - \frac{d_C^2(\mathbf{z}_t, \mathbf{z}_s)}{\varepsilon} \right\}
  \]

  - Approximates the Euclidean distances between the underlying process
  
  - Invariant to the measurement modality and resilient to noise

- The corresponding Laplace operator \( \mathbf{L} \) can recover the underlying process

- Compute the eigenvalues \( \{ \lambda_i \}_{i=1}^N \) and eigenvectors \( \{ \varphi_i \}_{i=1}^N \) of \( \mathbf{L} \)

- The leading eigenvectors represent the underlying process
Molecular Dynamics

[Dsilva, Talmon, Rabin, Coifman & Kevrekidis, 13’]

- Task: track Alanine Dipeptide in water from partial observations
• **Molecular Dynamics**

[Dsilva, Talmon, Rabin, Coifman & Kevrekidis, 13’]

- 3 Snapshots of the true and reconstructed trajectories
  - Using a multiscale method (Laplacian Pyramid [Rabin & Coifman, SDM, 12’])
• **Applications**

• **Predicting Epileptic Seizures**

**Results:**

- **3D points** - the 3 leading eigenvectors (each point – an EEG time frame)
• Bayesian Filtering

How to do processing in the inferred parametric domain?
  
  – Combine the inferred geometry and the time series dynamics

Main idea:

  – Define a pseudo-likelihood function based on the inferred intrinsic model
  – Empirically define a prior function based on past observations
  – Combine the two using a Bayesian framework

Assumption:

  – Locally, the distribution in the embedded domain is a good approximation of the distribution in the underlying process original domain
  – The posterior pdf of the underlying process $p(\theta_t | \theta_{t-1}, z_t)$ can be estimated based on the embedding
• Pseudo-likelihood

\[ \Psi(z_t) | \theta_t \sim \mathcal{N}(\theta_t, C_{\theta,t}) \]

Inaccuracy of the representation
Explicit use of the chronological order to obtain empirical dynamical model

\[ \theta_t | \theta_{t-1} \sim \mathcal{N} \left( \theta^f_{t-1}, C^f_{\theta,t-1} \right) \]
Represent the posterior pdf by a set of (random) samples

\[ p(\theta_t | \theta_{t-1}, z_t) \]

- Let \( \{ \theta_t^{(k)} \}_{k=1}^{P} \) be a set of support samples ("particles") that characterizes the posterior pdf given the previous stage and the new measurement.
- Let \( \{ w_t^{(k)} \}_{k=1}^{P} \) be a set of weights associated with the particles.

\[ p(\theta_t | \theta_{t-1}, z_t) \approx \sum_{k=1}^{P} w_t^{(k)} \delta(\theta_t - \theta_t^{(k)}) \]
\textbullet{} Where the weights are denoted as

\[ w_{t}^{(k)} \triangleq p(\theta_{t}^{(k)} | \theta_{t-1}, z_{t}) \]

with \( \sum_{k=1}^{P} w_{t}^{(k)} = 1 \)

\textbullet{} By Bayes’ theorem

\[
\begin{align*}
  w_{t}^{(k)} & \propto p(\theta_{t}^{(k)} | \theta_{t-1})p(z_{t} | \theta_{t}^{(k)}) \\
\end{align*}
\]

where the densities can be estimated in the embedded domain.
• MMSE Estimator

For example, based on the estimated posterior pdf, the MMSE estimator of the factors at $t$ can be computed by

$$
\hat{\theta}_t = \mathbb{E}[\theta_t | \theta_{t-1}, z_t] = \int \theta_t p(\theta_t | \theta_{t-1}, z_t) d\theta_t
$$

$$
\approx \sum_{k=1}^{P} p(\theta_t^{(k)} | \theta_{t-1}, z_t) \theta_t^{(k)} = \sum_{k=1}^{P} w_t^{(k)} \theta_t^{(k)}
$$

requires few initial values of the original factors for alignment.
• Biomedical Imaging

- Imaging model - consider a 2D shape measured by a 1D linear sensor array
  - Rigid biological material that vibrates over time
  - Emission of radiation or light
  - Noisy measurements of the instantaneous amount of radiation that travelled through the object
• **Objective:** to track the object based on the measurements

We simulated a diffusion process and output signals of 15 sensors with Gaussian noise:

\[ z_t^i = f(p_i - \theta_t) + v_t^i \]

• **Note:** the simulated model was not used for the inference and the tracking
• Tracking the center position of the shape:
  
  – The yellow curve is the true position of the center
  
  – The vertical gray strips represent the posterior pdf estimate
  
  – The solid black curve is the expected value (MMSE estimator)
• **Summary**

- The notion of *empirical intrinsic modeling*
  - Empirical geometry of local distributions

- *Nonlinear processing framework* in the low dimensional intrinsic domain

- Used in a wide variety of applications
  - Nonlinear problems without existing definitive models
  - In particular, *biomedical imaging*
    (based, for example, on photon counter sensors)
THANK YOU