Parametrization of Linear Systems Using Diffusion Kernels

paper by: Ronen Talmon, Dan Kushnir, Ronald R. Coifman, Israel Cohen and Sharon Gannot

Eli Kravchik

June 24, 2012
A system can be viewed as a "black box" controlled by a number of parameters.

We assume the output of the system is determined by these parameters.

A data driven supervised approach is proposed to recover parameters and characterize system.
Notation

- \(x\) - input signal
- \(y\) - measurement
- \(c\) - observation, obtained by
  \[
  c_y(\tau) = \mathbb{E}y(t)y(t + \tau) = h_\theta(\tau) * h_\theta(-\tau) * c_x(\tau)
  \]

  when the last equality holds in the linear case

- \(\bar{\Theta}, \Theta \subset \mathbb{R}^d\) - Training and test parameter sets
- \(\bar{\Gamma}, \Gamma \subset \mathbb{R}^D\) - Training and test observation sets
Ito Process Assumption

We assume that the parameters $\theta_i$ behave like as Ito process

$$d\theta^{(j)} = a^{(j)}(\theta^{(j)})dt + b^{(j)}(\theta^{(j)})dw$$

where $j = 1, \ldots, d$
Objective

- For $\theta_i \in \Theta, c_i \in \Gamma$ when $c_i = r(\theta_i)$ we would like to recover the parameter $\theta_i$ without prior knowledge of the function $r : \Theta \rightarrow \Gamma$ and without estimating it.

- We want to uncover the hidden geometry of the parameter space $\Theta$ from the observation space $\Gamma$ using knowledge from training parameter set $\overline{\Theta}$ and the corresponding training observation set $\overline{\Gamma}$.
Diffusion Maps

- Given a set $X \in \mathbb{R}^N$ we can use diffusion maps to capture its intrinsic geometrical structure - re-parameterizing it into a low dimensional manifold.
- The diffusion kernel is used to compute similarity Markov matrix: $W$
- It can be shown (under certain conditions) that eigenvectors $\{\phi_j\}$ of $W$ converge to eigenfunctions of the Laplace-Beltrami operator (FP operator for non uniform data distribution).
- $\{\phi_j\}$ give us the needed low dimensional embedding
Problem

- We need distance in the parametric space but have only the observation $c_i$

- We cannot directly imply the geometry of $\Theta$ from the geometry of $\Gamma$
From Distance in the Observation Space to Distance in the Parameter Space

From Taylor series expansion we get (when \( z = \frac{c_1 + c_2}{2} \)):

\[
\begin{align*}
\theta_1^{(i)} &= g(c_1)^{(i)} = \gamma^{(i)} + \frac{1}{2} \sum_{k=1}^{D} g_k^i (z)(c_1^{(k)} - c_2^{(k)}) \\
&+ \frac{1}{8} \sum_{k,l=1}^{D} g_{kl}^i (c_1^{(k)} - c_2^{(k)})(c_1^{(l)} - c_2^{(l)}) + O \| c_1 - c_2 \|_3
\end{align*}
\]

and in the same manner:

\[
\begin{align*}
\theta_2^{(i)} &= g(c_2)^{(i)} = \gamma^{(i)} + \frac{1}{2} \sum_{k=1}^{D} g_k^i (z)(c_2^{(k)} - c_1^{(k)}) \\
&+ \frac{1}{8} \sum_{k,l=1}^{D} g_{kl}^i (c_2^{(k)} - c_1^{(k)})(c_2^{(l)} - c_1^{(l)}) + O \| c_2 - c_1 \|_3
\end{align*}
\]
From Distance in the Observation Space to Distance in the Parameter Space

Leading to:

\[ \| \theta_1 - \theta_2 \|^2 = (c_1 - c_2)^T 2[J J^T (c_1) + J J^T (c_2)]^{-1} (c_1 - c_2) + O \| c_1 - c_2 \|^4 \]
For each sample $\bar{c}_i$ at the training set we compute the local covariance matrix $\Sigma(\bar{c}_i)$ by:

$$\Sigma(\bar{c}_i) = \frac{1}{L} \sum_{j=1}^{L} (c_{ij} - \mu_i)(c_{ij} - \mu_i)^T$$

where $\mu_i = \frac{1}{L} \sum_{j=1}^{L} c_{ij}$ and $\{c_{ij}\}_{j=1}^{L}$ are a cloud of observations around $\bar{c}_i$ corresponding to a cloud of small perturbations around the training parameter $\bar{\theta}_i$ for which $\bar{c}_i = r(\bar{\theta}_i)$.
Using Ito lemma we get:

$$dc^{(j)} = \sum_{i=1}^{d} \frac{1}{2} (b^{(i)})^2 r_{ii}^{j} + a^{(i)} r_{i}^{j} dt + \sum_{i=1}^{d} b^{(i)} r_{i}^{j} dw$$

Where $r : \Theta \rightarrow \Gamma$ is the (unknown) mapping from the parameters to the observations.

After normalizing and writing in matrix form we get the needed connection:

$$J(c)J^T(c) = \Sigma(c)$$
We define a diffusion kernel by:

\[
W^{(kj)} = \frac{\pi}{\sqrt{\det(J^T J(z))}} \exp \left( -\frac{(\vec{c}_j - \vec{c}_k)^T (J J^T (\vec{c}_j) + J J^T (\vec{c}_k))^{-1} (\vec{c}_j - \vec{c}_k)}{\epsilon} \right)
\]

and the affinity matrix:

\[
A(kj) = \exp \left( \frac{\|J^{-1} (\vec{c}_j) (\vec{c}_j - c_k)\|^2}{\epsilon} \right)
\]
Substituting $J(c)J^T(c) = \Sigma(c)$ into the diffusion kernel will give us:

$$W^{(kj)} = \frac{\pi}{\sqrt{\det(J^T J(z))}} \exp \left( -\frac{(\bar{c}_j - \bar{c}_k)(\Sigma(\bar{c}_j) + \Sigma(\bar{c}_k))^{-1}(\bar{c}_j - \bar{c}_k)^T}{\epsilon} \right)$$

and:

$$A(kj) = \exp \left( \frac{(\bar{c}_j - c_k)^T \Sigma^{-1}(\bar{c}_j)(\bar{c}_j - c_k)}{\epsilon} \right)$$
Diffusion Kernel Computation

It can be shown that:

\[ W = S^{-\frac{1}{2}} A^T A S^{-\frac{1}{2}} \]

meaning that a distance in the parameter space between two training set points is determined by how they are seen by the test set.
Let \( \{\lambda_j\}_{j=1}^{m} \) and \( \{\phi_j\}_{j=1}^{d} \in \mathbb{R}^d \) be the leading \( d \) eigenvalues and eigenvectors of \( W \). Then the map \( \Phi_d : \bar{\Gamma} \rightarrow \mathbb{R}^d \)

\[
\Phi_d(\bar{c}_i) = [\phi_1(\bar{c}_i), \ldots, \phi_d(\bar{c}_i)]^T
\]

will be an embedding of the training set parameters into \( \mathbb{R}^d \).
Low Dimensional Embedding

We can compute the leading $d$ eigenvectors of $\tilde{A}\tilde{A}^T \{\psi_j\}_{j=1}^d$ by:

$$\psi_j = \frac{1}{\lambda_j} \tilde{A}\phi_j$$

and get a re-parameterization of the test observations into $\mathbb{R}^d$ by the map $\Psi_d : \Gamma \rightarrow \mathbb{R}^d$

$$\Psi_d(c_i) = [\psi_1(c_i), \ldots, \psi_d(c_i)]^T$$
Using the embedding $\Psi_d(c_i)$ we can interpolate the parameter $\theta_i$ corresponding to the observation $c_i$:

$$\gamma_j(c_i) = \frac{exp\left(\frac{-\|\Psi_d(c_i) - \Psi_d(\bar{c}_j)\|^2}{\sigma_j}\right)}{\sum_{\Psi_d(\bar{c}_k) \in N_i} exp\left(\frac{-\|\Psi_d(c_i) - \Psi_d(\bar{c}_k)\|^2}{\sigma_j}\right)}$$

$$\hat{\theta}_i = \sum_{\Psi_d(\bar{c}_j) \in N_i} \gamma_j(c_i) \bar{\theta}_j$$
The AR model is presented by the connection:

\[ y(n) = - \sum_{i=1}^{p} a_i y(n - i) + \epsilon(n) \]

Given \( p = 2 \) and two poles \( \theta_1 \sim U[0.2, 0.8], \theta_2 \sim U[-0.8, -0.2] \)
for which \( a_1 = -(\theta_1 + \theta_2) \) and \( a_2 = \theta_1 \theta_2 \) we want to approximate the poles \( \theta_1, \theta_2 \) from measured signal \( y(n) \) by the method described in the paper.
Mean error was: \( \text{err} = 0.0376 \).
AR Model Pole Estimation

Scatter plots of embedded samples on the leading eigenvectors of diffusion kernel:
Noisy Measurements

![Graph showing error vs noise](image-url)

The graph illustrates the relationship between error and noise standard deviation (noise std). As the noise std increases, the error also increases, indicating a higher level of uncertainty in the measurements.
Training Set Size

![Graph showing error vs training set size](image-url)
Error Analysis

$$e_i = \|\theta_i - \hat{\theta}_i\| = \|\theta_i - \sum_{\Psi_d(\bar{c}_j) \in N_i} \gamma_j(c_i)\bar{\theta}_j\|$$

$$\leq \sum_{\Psi_d(\bar{c}_j) \in N_i} \gamma_j(c_i)\|\theta_i - \bar{\theta}_j\| \leq N \max_{\Psi_d(\bar{c}_j) \in N_i} \|\theta_i - \bar{\theta}_j\|$$

The error depends on how well the diffusion kernel can approximate the geometry of the parameter space.
Future Work

- Testing method on other models (non linear) and distributions.
- Mathematical connection between training set and error
- Response to different input signals
- Temporal approach
References

- Parametrization of Linear Systems Using Diffusion Kernels by Ronen Talmon Dan Kushnir, Ronald R. Coifman, Israel Cohen and Sharon Gannot

- Anisotropic Diffusion on Sub-Manifolds with Application to Earth Structure Classification by D. Kushniry, A. Haddady, R. R. Coifman

- Non-linear independent component analysis with diffusion maps by Amit Singer and Ronald R. Coifman

- Spectral independent component analysis by A. Singer
- Differential Stochastic Sensing: Intrinsic Modeling of Random Time Series with Applications to Nonlinear Tracking by Ronen Talmon and Ronald R. Coifman
- Supervised Graph-based Processing for Sequential Transient Interference Suppression by Ronen Talmon, Israel Cohen Sharon Gannot and Ronald R. Coifman
- Diffusion maps by Ronald R. Coifman and Stephane Lafon
- Acknowledgment: Dr Ronen Talmon
Thank You for Your Time

Questions?