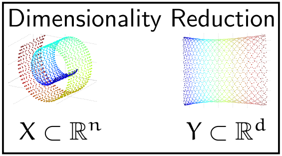
Dimensionality Reduction and Persistent Homology in Signal Analysis

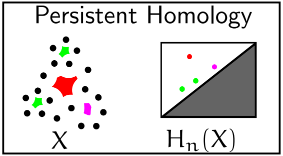
Mijail Guillemard

SeisMath
Mathematical Models in Seismology
August 27 - September 7, 2012
L'Aquila, Italy

Overview

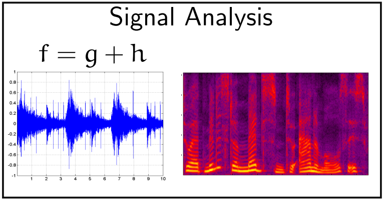


PCA, MDS, Kernel Methods,
Isomap, Laplacian Eigenmaps



Simplicial Homology, category theory

$$X = \{x_i\}_{i=1}^{m} \subset \mathbb{R}^n$$



Time Frequency Analysis, Group Representations
Signal Separation

Dimensionality Reduction and Manifold Learning

Point Cloud Data $X = \{x_i\}_{i=1}^m \subset \mathcal{M} \subset \mathbb{R}^n$

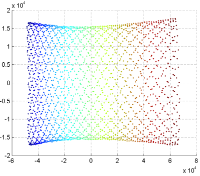
Hypothesis:

- $\mathcal{M}$ manifold, topological space (simplicial complex)
- $\dim(\mathcal{M}) = p \ll n$
- $\mathbb{R}^d \supset \Omega \xrightarrow{\mathcal{A}} \mathcal{M} \subset \mathbb{R}^n$ homeomorphism, $d < n$.

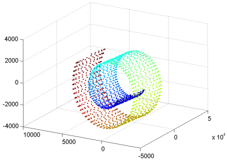
Objectives:

- Construct $Y = \{y_i\}_{i=1}^m \subset \Omega' \subset \mathbb{R}^d$, $d < n$
- $\mathbb{R}^n \supset \mathcal{M} \xrightarrow{P} \Omega' \subset \mathbb{R}^d$ homeomorphism (diffeomorphism)

$\mathrm{Y} \subset \Omega' \subset \mathbb{R}^2$



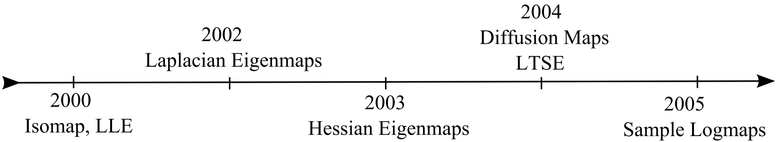
$\mathrm{X} \subset \mathcal{M} \subset \mathbb{R}^{1024}$



$$\begin{tikzcd}
\mathbb{R}^d \supset \Omega \arrow[r, "\mathcal{A}"] & \mathcal{M} \subset \mathbb{R}^n \arrow[dl, "P"] \\
\mathbb{R}^d \supset \Omega' &
\end{tikzcd}$$

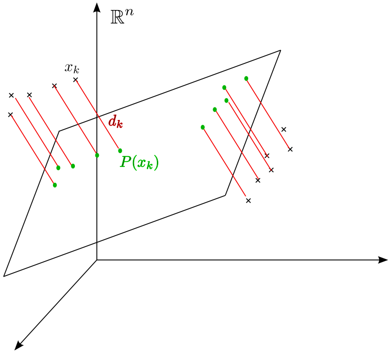
Dimensional Reduction Techniques

- PCA
- Linear Discriminant Analysis (Fisher)
- Generalized Discriminant Analysis
- Multidimensional Scaling
- Isomap
- Supervised Isomap



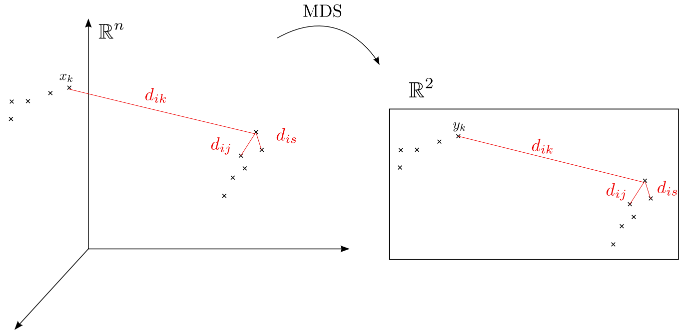
Principal Component Analysis (PCA)

- Matrix data: $X = (x_1 \dots x_m) \in \mathbb{R}^{n \times m}$
- Problem: find projection $P : \mathbb{R}^n \to \mathbb{R}^3$ with:
- $err(P, X) = \sum_k ||x_k - P(x_k)||^2$ minimum
- $\mathrm{var}(P(X)) = \sum_k ||P(x_k)||^2$ maximum
- "Maximum" Eigenvectors of the covariance matrix XX^t
- SVD of X



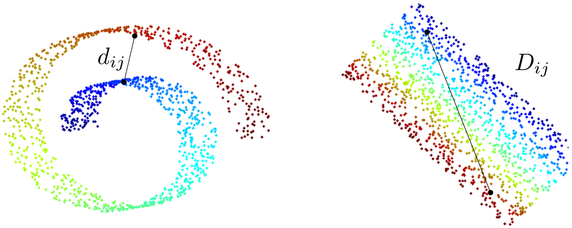
Multidimensional Scaling (MDS)

- Matrix data: $X = (x_1 \dots x_m) \in \mathbb{R}^{n \times m}$
- Problem: find a $Y = (y_1 \dots y_m) \in \mathbb{R}^{2 \times m}$ with:
- $err(Y, X) = \sum_k (d_{ij} - ||y_i - y_j||)^2$ is minimum
- $d_{ij} = ||x_i - x_j||$.



Isomap

- Geometric nonlinearities in the manifold
- Build a distance using the right geometry
- Apply MDS
- If you have the information, use supervised Isomap.



$$D_{ij} > d_{ij}$$

Example of a Functional Cloud for Speech Signals

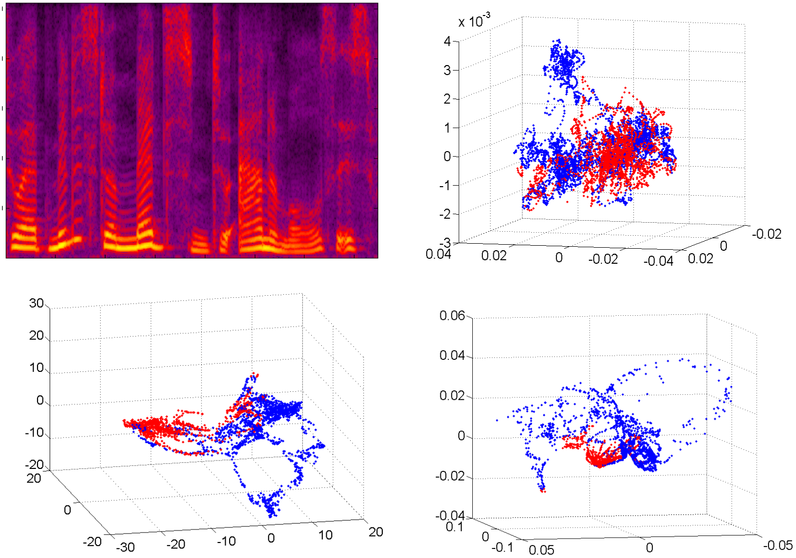
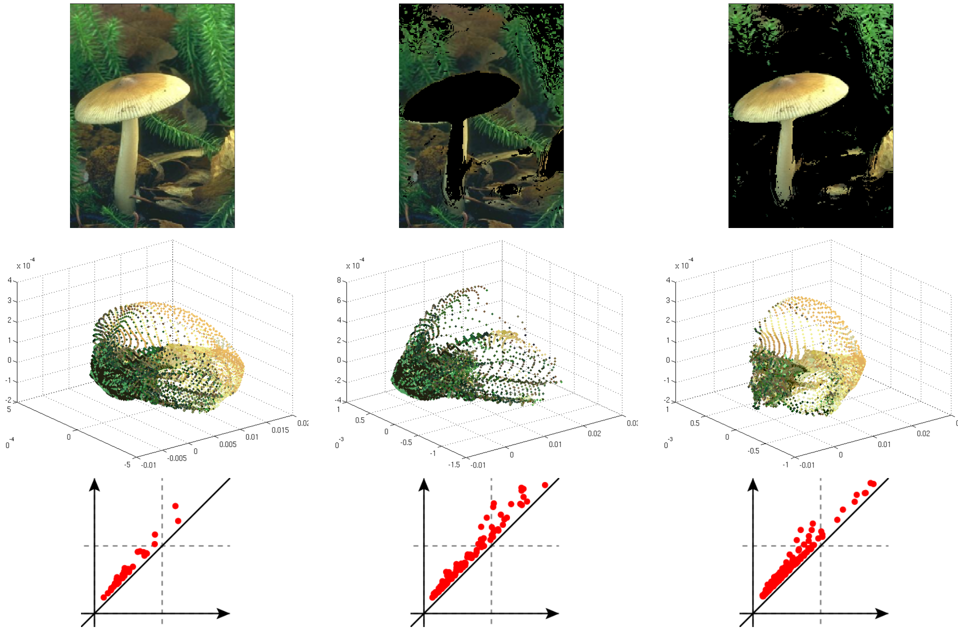
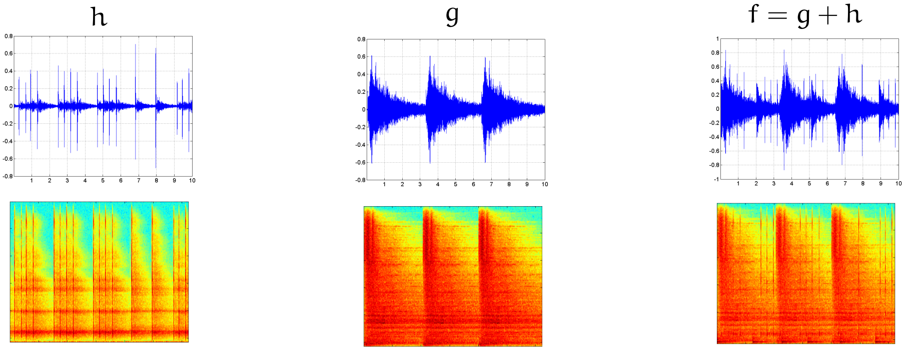
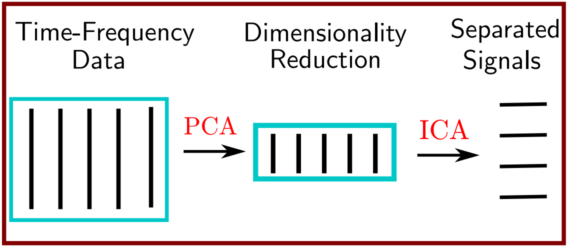


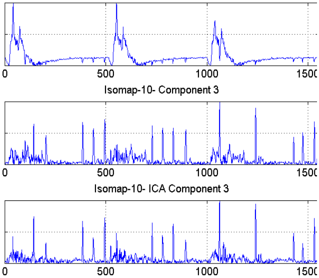
Image Processing Example



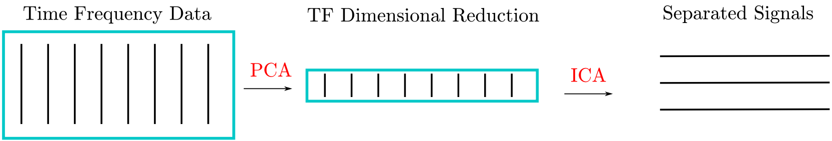
Signal Separation and Detection

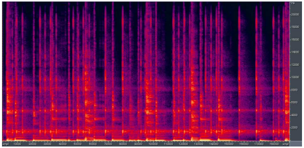




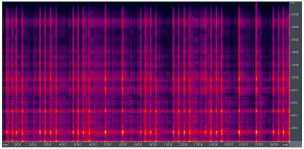


Mono Signal Source Separation

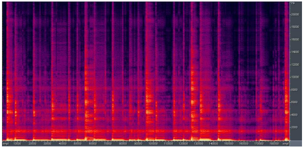




Castanets + Drum



Castanets

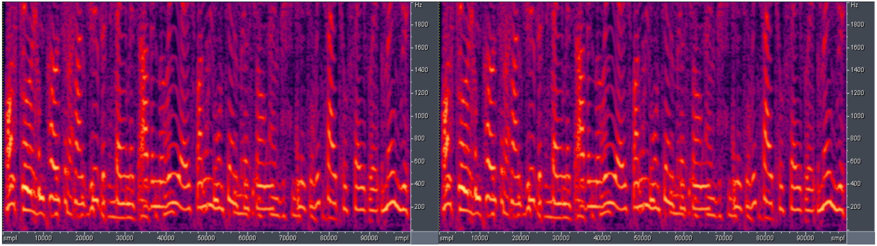


Drum

Stereo Source Separation

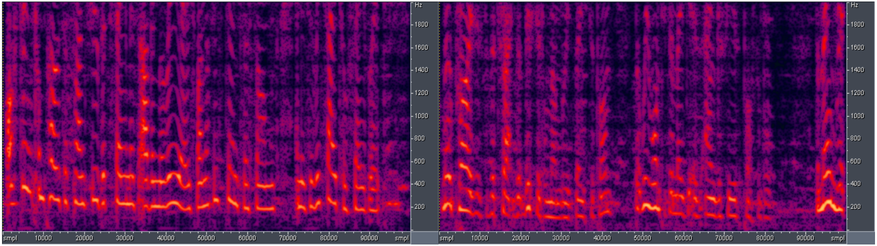
- N signal sources $s_j, j = 1, \cdots, N$ are mixed
- A stereo signal (x_1, x_2) is recorded with two microphones:
- $x_1(n) = \sum_{j=1}^{N} pl_j s_j(n)$
- $x_2(n) = \sum_{j=1}^{N} pr_j s_j(n)$
- pl_j is the left panning of the signal source j
- pr_j is the right panning of the signal source j
- D. Barry and B. Lawlor: Source Separation: Azimuth Discrimiation and Resynthesis

Stereo Source Separation

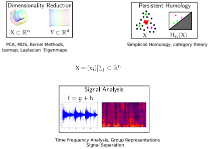


Separating the voice of two speakers: Each time-frequency atom, representing harmonic or noise components, is split in two parts.

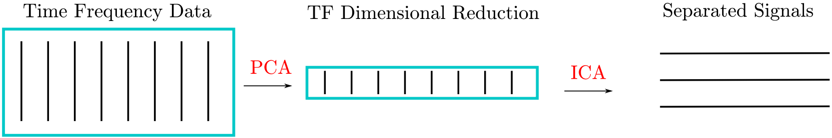
Stereo Source Separation

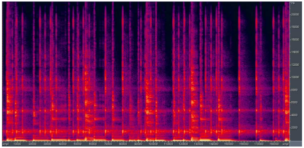


Separating the voice of two speakers: Each time-frequency atom, representing harmonic or noise components, is split in two parts.

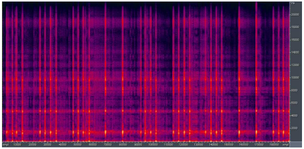

Dimensionality Reduction
X ⊂ ℝ^n
Y ⊂ ℝ^d
PCA, MDS, Kernel Methods,
Isomap, Laplacian Eigenmaps
Persistent Homology
X
H_n(X)
Simplicial Homology, category theory
X = {x_i}_{i=1}^m ⊂ ℝ^n
Signal Analysis
f = g + h
Time Frequency Analysis, Group Representations
Signal Separation

Mono Signal Source Separation

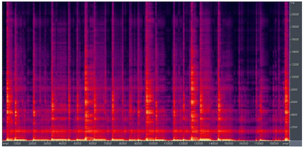




Castanets + Drum



Castanets



Drum