Analysis of High-Dimensional Signal Data by Manifold Learning and Convolution Transforms

Mijail Guillemard and Armin Iske
University of Hamburg

Alba di Canazei, 6-9 September 2010

Outline

Application from Neuro and Bioscience

- Electromyogram (EMG) Signal Analysis

Objectives and Problem Formulation

- Manifold Learning and Convolution Transforms

Some more Background

- Dimensionality Reduction: PCA, MDS, and Isomap
- Differential Geometry: Curvature Analysis

Numerical Examples

- Parameterization of Scale- and Frequency-Modulated Signals
- Manifold Evolution under Wave Equation
- Geometrical and Topological Distortions through Convolutions

EMG Signal Analysis in Neuro and Bioscience

- An EMG signal is an electrical measurement combining multiple action potentials propagating along motor neural cells.
- EMG signal analysis provides ...
  ... physiological information about muscle and nerve interactions.
- Applications:
  diagnosis of neural diseases, athletic performance analysis.

Goal: Develop fast and accurate methods to EMG signal analysis.

Tools:

- Dimensionality Reduction;
- Manifold Learning.

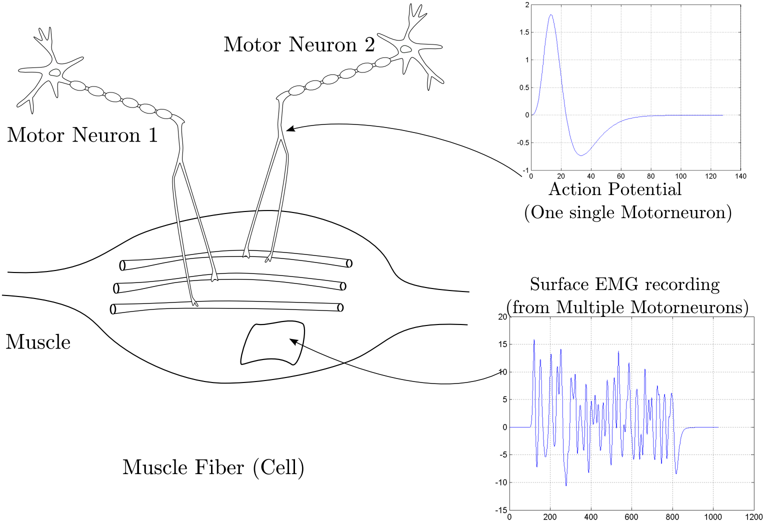
Peggy busy weight lifting.

Collaborators:

Deutscher Olympia-Stützpunkt Rudern, Hamburg

Prof. Klaus Mattes (Bewegungs- und Trainingswissenschaft, UHH)

Action Potential and Surface EMG



Manifold Learning by Dimensionality Reduction.

Input data: $X = \{x_1, \ldots, x_m\} \subset \mathcal{M} \subset \mathbb{R}^n$;

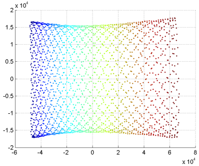
Hypothesis:

- $Y = \{y_1, \ldots, y_m\} \subset \Omega \subset \mathbb{R}^d$, $d \ll n$;
- nonlinear embedding map $\mathcal{A} : \Omega \to \mathbb{R}^n$, $X = \mathcal{A}(Y)$;

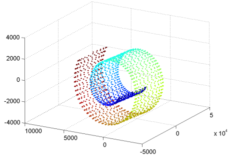
Task: Recover Y (and Ω) from X.

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

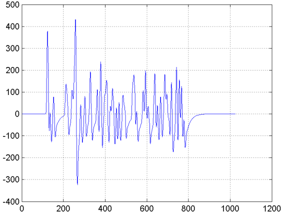
$Y \subset \Omega \subset \mathbb{R}^2$



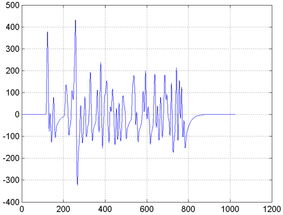
$X \subset \mathbb{R}^{1024}$



Objectives and Method Description



$\vdots$



$$X = \{x_1, \ldots, x_m\} \subset \mathcal{M} \subset \mathbb{R}^n$$

$$\begin{tikzcd}
\Omega \subset \mathbb{R}^d \arrow[r, "\mathcal{A}"] & \mathcal{M} \subset \mathbb{R}^n \arrow[d, "T"] \\
\Omega' \subset \mathbb{R}^d & \mathcal{M}_T \subset \mathbb{R}^n \arrow[l, "\mathcal{P}"']
\end{tikzcd}$$

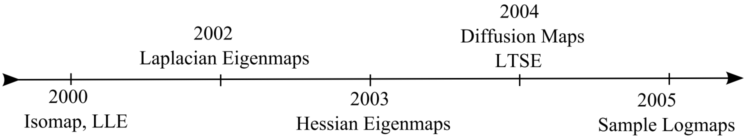
T Signal Transformation (Wavelets, Fourier,...)

$\mathcal{P}$ Dimensionality Reduction

Manifold Learning Techniques

Geometry-based Dimensionality Reduction Methods.

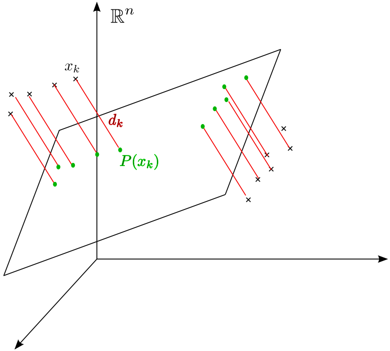
- Principal Component Analysis (PCA)
- Multidimensional Scaling (MDS)
- Isomap
- Supervised Isomap
- Local Tangent Space Alignment (LTSA)
- Riemannian Normal Coordinates (RNC, 2005)



Principal Component Analysis (PCA)

Problem:
Given points $X = \{x_k\}_{k=1}^m \subset \mathbb{R}^n$, find closest hyperplane $H \subset \mathbb{R}^n$ to X, i.e., find orthogonal projection $P : \mathbb{R}^n \to \mathbb{R}^n$, with $\operatorname{rank}(P) = p < n$, minimizing

$$d(P, X) = \sum_{k=1}^{m} \|x_k - P(x_k)\|^2$$



Principal Component Analysis (PCA)

Theorem: Let $X = (x_1, \dots, x_m) \in \mathbb{R}^{n \times m}$ be scattered with zero mean,

$$\frac{1}{m} \sum_{k=1}^{m} x_k = 0.$$

Then, for an orthogonal projection $P : \mathbb{R}^n \to \mathbb{R}^n$, with $\mathrm{rank}(P) = p < n$, are equivalent:

(a) The projection P minimizes the distance

$$d(P, X) = \sum_{k=1}^{m} \|x_k - P(x_k)\|^2.$$

(b) The projection P maximizes the variance

$$\mathrm{var}(P(X)) = \sum_{k=1}^{m} \|P(x_k)\|^2.$$

(c) The matrix representation of the projection P is spanned by p eigenvectors of the p largest eigenvalues of the covariance matrix $XX^T \in \mathbb{R}^{n \times n}$.

Proof: By the Pythagoras Theorem, we have

$$\|x_k\|^2 = \|x_k - P(x_k)\|^2 + \|P(x_k)\|^2 \qquad \text{for } k = 1, \ldots, m$$

and so

$$d(P, X) = \sum_{k=1}^{m} \|x_k\|^2 - \mathrm{var}(P(X)).$$

Therefore, minimizing $d(P, X)$ is equivalent to maximizing $\mathrm{var}(P(X))$, i.e., characterizations (a) and (b) are equivalent.
To study property (c), let $v_1, \ldots, v_p \subset \mathbb{R}^n$ denote an orthonormal set of vectors, such that

$$P(x_k) = \sum_{i=1}^{p} < v_i, x_k > v_i$$

and therefore

$$\mathrm{var}(P(X)) = \sum_{k=1}^{m} \|P(x_k)\|^2 = \sum_{k=1}^{m} \sum_{i=1}^{p} | < v_i, x_k > |^2.$$

Principal Component Analysis (PCA)

Now,

$$\begin{aligned}\sum_{k=1}^{m} | < v_i, x_k > |^2 &= \sum_{k=1}^{m} < v_i, x_k >< x_k, v_i > \\ &= \left\langle v_i, \sum_{k=1}^{m} x_k < x_k, v_i > \right\rangle \\ &= \left\langle v_i, \sum_{k=1}^{m} x_k \sum_{j=1}^{n} x_{jk} v_{ji} \right\rangle\end{aligned}$$

For the covariance matrix

$$S = XX^T = \left(\sum_{k=1}^{m} x_{\ell k} x_{jk} \right)_{1 \leq \ell, j \leq n}$$

we obtain

$$(Sv_i)_\ell = \sum_{j=1}^{n} \sum_{k=1}^{m} x_{\ell k} x_{jk} v_{ji} = \sum_{k=1}^{m} x_{\ell k} \sum_{j=1}^{n} x_{jk} v_{ji}$$

Principal Component Analysis (PCA)

Therefore,

$$\sum_{k=1}^{m} | < v_i, x_k > |^2 = < v_i, Sv_i >$$

and so

$$\text{var}(P(X)) = \sum_{k=1}^{m} \|P(x_k)\|^2 = \sum_{i=1}^{p} < v_i, Sv_i >$$

Now, maximizing $\text{var}(P(X))$ is equivalent to select p eigenvectors $v_1, \ldots, v_p$ corresponding to the p *largest* eigenvalues of S. □

The latter observation relies on the following

Lemma: Let $S \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Then, the function

$$W(v) = < v, Sv > \qquad \text{for } v \in \mathbb{R}^n \text{ with } \|v\| = 1$$

attains its maximum at an eigenvector of S corresponding to the *largest* eigenvalue of S.

Proof: Exercise.

Conclusion.

In order to compute P minimizing $d(P, X)$, perform the following steps

(a) Compute the singular value decomposition of X, so that $X = UDV^T$, with a diagonal matrix D containing the singular values of X, and U, V unitary matrices.

(b) So obtain

$$XX^T = U(DD^T)U^T$$

the eigendecomposition of the covariance matrix XX^T.

(c) Take the p (orthonormal) eigenvectors $v_1, \dots, v_p$ in U corresponding to the p largest singular values in DD^T, and so obtain the required projection

$$P(x) = \sum_{i=1}^{p} < x, v_i > v_i.$$

$\square$

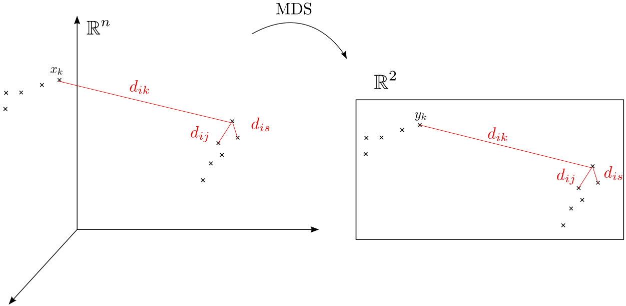
Multidimensional Scaling (MDS)

Problem (MDS):
Given distance matrix

$$D_X = (\|x_i - x_j\|^2)_{1 \le i,j \le m} \in \mathbb{R}^{m \times m}$$

of $X = (x_1 \ldots x_m) \in \mathbb{R}^{n \times m}$, find $Y = (y_1 \ldots y_m) \in \mathbb{R}^{p \times m}$ minimizing

$$d(Y, D) = \sum_{i,j=1}^{m} (d_{ij} - \|y_i - y_j\|^2).$$



Solution to Multidimensional Scaling Problem.

Theorem: Let $D = (\|x_i - x_j\|^2)_{1 \leq i,j \leq m} \in \mathbb{R}^{m \times m}$ be a given distance matrix of $X = (x_1, \ldots, x_m) \in \mathbb{R}^{n \times m}$ with

$$\frac{1}{m} \sum_{i=1}^{m} x_i = 0.$$

Then, the matrix X can (up to orthogonal transformation) be recovered from D by

$$Y = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_m}) U^T,$$

where $\lambda_1 \geq \ldots \geq \lambda_m \geq 0$ and $U \in \mathbb{R}^{m \times n}$ are the corresponding eigenvalues and eigenvectors of the matrix

$$XX^T = -\frac{1}{2} JDJ \quad \text{with } J = I - (1/m)ee^T$$

where $e = (1, \ldots, 1)^T \in \mathbb{R}^m$.

Multidimensional Scaling (MDS)

Proof (sketch): Note that

$$d_{ij} = \|x_i - x_j\|^2 = (x_i, x_i) - 2(x_i, x_j) + (x_j, x_j).$$

This implies the relation

$$D = Ae^T - 2XX^T + eA^T$$

for $A = [(x_1, x_1), \ldots, (x_m, x_m)]^T \in \mathbb{R}^m$. Now regard $J = I_m - (1/m)ee^T \in \mathbb{R}^{m \times m}$. Since $Je = 0$, the above relation implies

$$XX^T = -\frac{1}{2}JDJ.$$

Therefore, the eigendecomposition of $XX^T = U\mathrm{diag}(\lambda_1, \ldots, \lambda_m)U^T$ allows us to recover X, up to orthogonal transformation, by

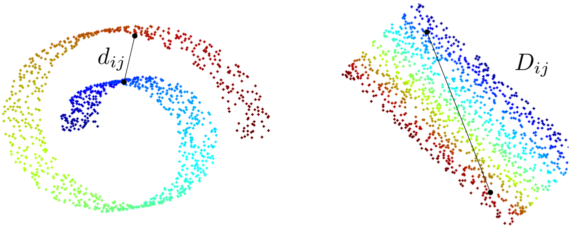
$$Y = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_m})U^T.$$

$\square$

From Multidimensional Scaling to Isomap.

Example: Regard the Swiss roll data, sampled from the surface

$$f(u,v) = (u\cos(u), v, u\sin(u)) \qquad \text{for } u \in [3\pi/2, 9\pi/2], v \in [0,1].$$



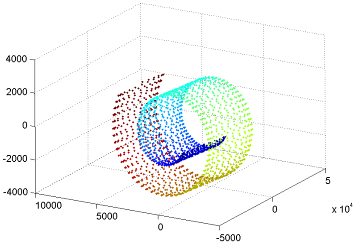
$$D_{ij} > d_{ij}$$

Isomap Strategy.

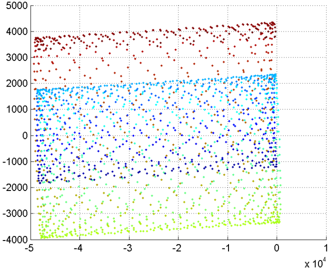
- **Neighbourhood graph construction.**
  Define a graph where each vertex is a data point, and each edge connects two points satisfying an ϵ-radius or k-nearest neighbour criterion.
- **Geodesic distance construction.**
  Compute the geodesic distance between each point pair (x, y) by using the length of shortest path from x to y in the graph.
- ***d*-dimensional embedding.**
  Use the geodesic distances for computing a d-dimensional embedding as in the MDS algorithm.

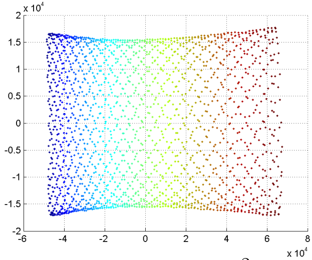
Dimensionality Reduction by Isomap

Swiss Roll Dataset $\mathbb{R}^3$

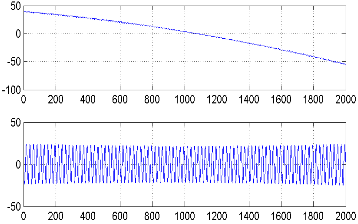


PCA projection $\mathbb{R}^2$





Isomap projection $\mathbb{R}^2$



Isomap projection: Eigenvectors

Curvature of Curves.

- For a curve $r : I \longrightarrow \mathbb{R}^n$ with arc-length parametrization:

$$s(t) = \int_a^t \|r'(x)\|\, dx$$

its curvature is

$$\kappa(s) = \|r''(s)\|$$

- For a curve $r : I \longrightarrow \mathbb{R}^n$ with arbitrary parametrization we have

$$\boxed{\kappa^2(r) = \frac{\|\ddot{r}\|^2\|\dot{r}\|^2 - \langle \ddot{r}, \dot{r}\rangle^2}{(\|\dot{r}\|^2)^3}}$$

Computation of the Scalar Curvature for Manifolds.

Metric Tensor : $g_{ij}(p) = \langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \rangle$

Christoffel symbols : $\Gamma_{ij}^{k} = \frac{1}{2} \sum_{\ell=1}^{m} \left(\frac{\partial g_{j\ell}}{\partial x_i} + \frac{\partial g_{i\ell}}{\partial x_j} - \frac{\partial g_{ij}}{\partial x_\ell} \right) g^{\ell k}$

Curvature Tensor : $R^{\ell}{}_{ijk} = \sum_{h=1}^{m} (\Gamma_{jk}^{h} \Gamma_{ih}^{\ell} - \Gamma_{ik}^{h} \Gamma_{jh}^{\ell}) + \frac{\partial \Gamma_{jk}^{\ell}}{\partial x_i} - \frac{\partial \Gamma_{ik}^{\ell}}{\partial x_j}$

Ricci Tensor : $R_{ijk\ell} = \sum_{h=1}^{m} R_{ijk}^{h} g_{\ell h}, \quad R_{ij} = \sum_{k,\ell=1}^{m} g^{k\ell} R_{kij\ell} = \sum_{k=1}^{m} R_{kij}^{k}$

Scalar Curvature : $\kappa = \sum_{i,j=1}^{m} g^{ij} R_{ij}$

Curvature Distortion and Convolution

- Regard the convolution map T on manifold $\mathcal{M}$,

- $\mathcal{M}_T = \{T(x), x \in \mathcal{M}\} \quad T(x) = x * h, \quad h = (h_1, \ldots, h_m)$

$$T = \begin{pmatrix} h_1 & 0 & \ldots & 0 \\ h_2 & h_1 & \ldots & 0 \\ h_3 & h_2 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ h_m & h_{m-1} & \ldots & h_1 \\ 0 & h_m & \ldots & h_2 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & h_m \end{pmatrix}$$

- Curvature (for curves): $\boxed{\kappa_T^2(r) = \frac{\|T\ddot{r}\|^2 \|T\dot{r}\|^2 - \langle T\ddot{r}, T\dot{r}\rangle^2}{(\|T\dot{r}\|^2)^3}}$

Modulation Maps and Modulation Manifolds

- A *modulation map* is defined for $\alpha = (\alpha_1, \ldots, \alpha_d) \in \Omega$, and $\{t_i\}_{i=1}^n \subset [0,1]$:

$$\mathcal{A} : \Omega \to \mathcal{M} \quad \rightsquigarrow \quad \mathcal{A}_\alpha(t_i) = \sum_{k=1}^{d} \phi_k(\alpha_k t_i) \text{ for } 1 \le i \le n,$$

  ($\Omega \subset \mathbb{R}^d$ and $\mathcal{M} \subset \mathbb{R}^n$ manifolds, $\dim(\Omega) = \dim(\mathcal{M}),\ d < n$).

- Example: *Scale modulation map*

$$\mathcal{A}_\alpha(t_i) = \sum_{k=1}^{d} \exp(\alpha_k(t_i) - b_k)^2 \qquad \text{for } 1 \le i \le n.$$

- Example: *Frequency modulation map*

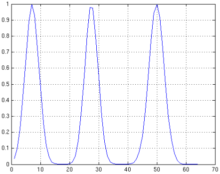
$$\mathcal{A}_\alpha(t_i) = \sum_{k=1}^{d} \sin(\alpha_k t_i + b_k) \qquad \text{for } 1 \le i \le n.$$

Numerical Example 1: Curvature Distortion

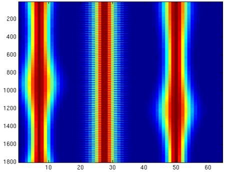
Low dimensional Parameterization of Scale Modulated Signals.

$$X = \left\{ f_{\alpha^t} = \sum_{i=1}^{3} e^{-\alpha_i(t)(\cdot - b_i)^2}, \alpha \in \Omega \right\}$$

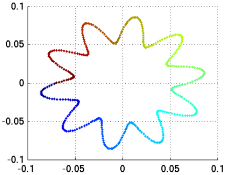
$$\Omega = \left\{ \alpha^t = (\alpha_1(t), \alpha_2(t), \alpha_3(t)), \quad t \in [t_0, t_1] \right\}$$



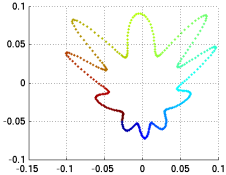
f_α



$X = \{f_\alpha, \alpha \in \Omega\}$



$\Omega \subset \mathbb{R}^3$

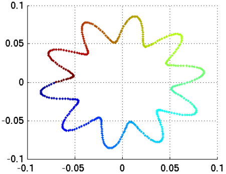


$X \subset \mathbb{R}^{64}$

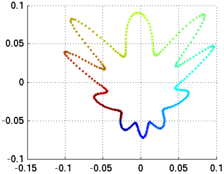
Low dimensional Parameterization of Scale Modulated Signals.

Multiresolution
$$\begin{tikzcd} V_j \arrow[r] \arrow[dr] & V_{j+1} \arrow[r] \arrow[dr] & V_{j+2} & \cdots \\ & W_{j+1} & W_{j+2} & \cdots \end{tikzcd}$$

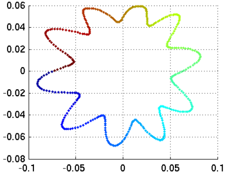
$$V_8 \oplus W_8 \oplus W_{16} \oplus W_{32} = V_{64}$$



$\Omega \subset \mathbb{R}^3$



$X \subset \mathbb{R}^{64}$

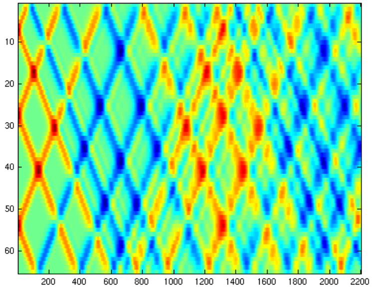


$T(X) \subset \mathbb{R}^{64}$

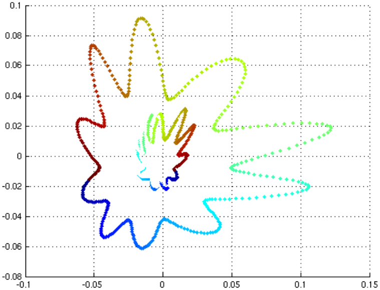
Manifold Evolution under a PDE

Wave Equation $$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} \qquad \text{(WE)}$$

$$U_t = \left\{ u_\alpha(t,x), \; u_\alpha \text{ solution of (WE) with initial condition } f_\alpha, \; \alpha \in \Omega_0 \right\}$$



$$\left\{ u_{\alpha_0}(t,x), (t,x) \in [t_0, t_1] \times [x_0, x_1] \right\}$$

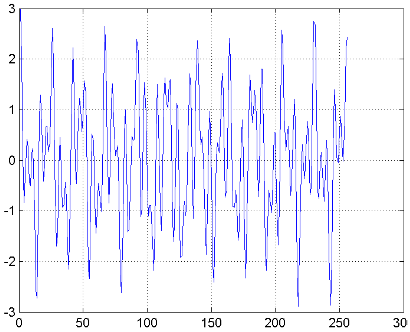


$X_0 \quad X_t$

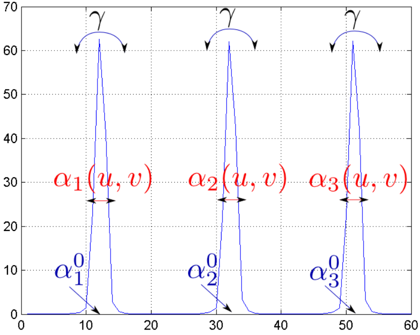
Frequency Modulation.

$$\boxed{\mathcal{A} : \Omega \subset \mathbb{R}^d \to \mathcal{M} \subset \mathbb{R}^n}$$

$$\begin{aligned} \alpha_1(u,v) &= (R + r\cos v)\cos u \\ \alpha_2(u,v) &= (R + r\cos v)\sin u \quad \xrightarrow{\mathcal{A}} \quad \mathcal{A}_\alpha(t_i) = \textstyle\sum_{k=1}^{3} \sin((\alpha_k^0 + \gamma\alpha_k)t_i) \\ \alpha_3(u,v) &= r\sin v \end{aligned}$$



$\mathcal{A}_\alpha(t)$

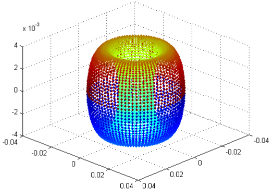


Frequency Content of $\mathcal{A}_\alpha(t)$

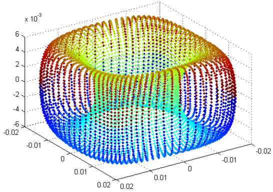
Torus Deformation.

$\mathcal{A} : \Omega \subset \mathbb{R}^d \to \mathcal{M} \subset \mathbb{R}^n$ $\qquad$ $\mathcal{P} : \mathcal{M} \subset \mathbb{R}^n \to \Omega' \subset \mathbb{R}^3$

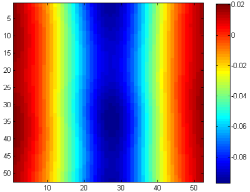
Modulation Map $\qquad$ PCA 3D Projection : $\mathcal{P}(\mathcal{M}) = \Omega'$



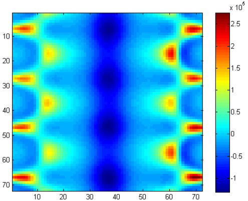
Torus $\Omega \subset \mathbb{R}^3$



$\mathcal{P}(\mathcal{M})$: PCA 3D Projection of $\mathcal{M} \subset \mathbb{R}^{256}$



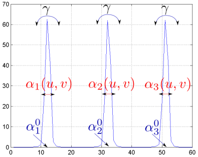
Scalar Curvature of $\mathcal{M} \subset \mathbb{R}^{256}$

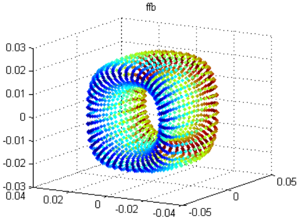


Scalar Curvature of $\mathcal{P}(\mathcal{M}) \subset \mathbb{R}^3$

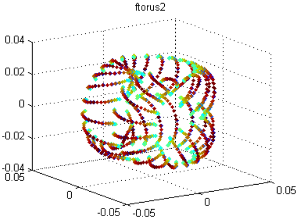
PCA Distortion for high frequency bandwidths γ.

$$\mathcal{A}_\alpha(t_i) = \sum_{k=1}^{3} \sin((\alpha_k^0 + \gamma\alpha_k)t_i)$$





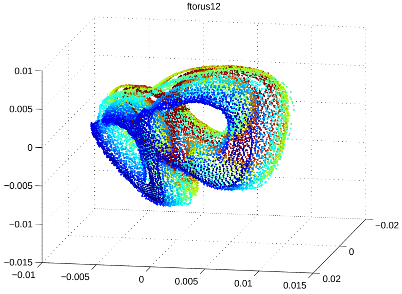
Isomap 3d Projection of $\mathcal{M}$

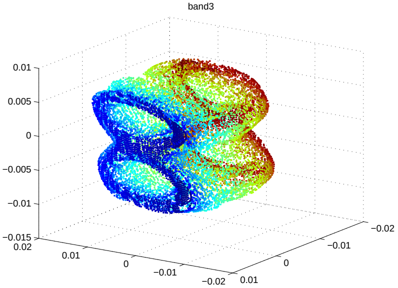


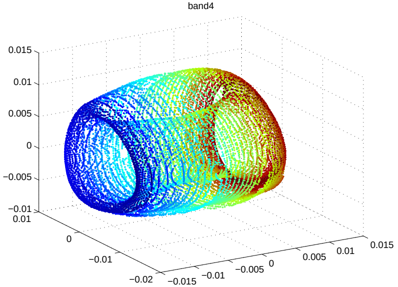
PCA 3d Projection of $\mathcal{M}$

Example 3: Topological Distortion

Torus (Genus 1 and Genus 2).







Relevant Literature.

- M. Guillemard and A. Iske (2010)
  Curvature Analysis of Frequency Modulated Manifolds in Dimensionality Reduction. *Calcolo*.
  Published online, 21st Sept. 2010, DOI: 10.1007/s10092-010-0031-8.
- M. Guillemard and A. Iske (2009)
  Analysis of High-Dimensional Signal Data by Manifold Learning and Convolutions. In: *Sampling Theory and Applications (SampTA'09)*, L. Fesquet and B. Torrésani (eds.), Marseille, May 2009, 287–290.
- Further up-to-date material, including slides and code, are accessible through our homepages
  `http://www.math.uni-hamburg.de/home/guillemard/` and
  `http://www.math.uni-hamburg.de/home/iske/`.