IT4I

Manifolds as a Useful Data Structure

Irina Perfilieva

University of Ostrava, Centre of Excellence IT4Innovations
Institute for Research and Applications of Fuzzy Modeling
30. dubna 22, Ostrava,
Czech Republic

irina.perfilieva@osu.cz

Seminar, Praha 24-03-2022

EVROPSKÁ UNIE
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ
INVESTICE DO VAŠÍ BUDOUCNOSTI

OP Výzkum
a vývoj pro inovace

IRAFM

Outline

Motivation

C. Anderson, "The End of Theory"

1

- "Traditional scientific method based on hypotheses would become obsolete".
- "No more theories or hypotheses, no more discussions about whether the experimental results refute or confirm the original hypotheses".
  - "In this new era, sophisticated algorithms and statistical tools are needed to sift through vast amounts of data and find information that can be turned into knowledge".

[1] *http://archive.wired.com)/science/discovery/journal/*

Data-driven versus hypothesis-driven research

Question 1

Is data-driven research a genuine way to gain knowledge, or is it primarily a tool for identifying potentially useful information?

Question 2

If we consider Newtonian models based on rough approximations to reality that are wrong at the atomic level, what do we conclude about their usefulness?

Inverse Problems

1st Answer

- The relationship between data and its model is analyzed using inverse problems
- They tell us about parameters that we cannot directly observe ... but
- They influence our judgment of the model validity in relation to the available data.
- The formally expressed relationship is

$$d_{obs} = F(p).$$

- The inverse problem is to determine the model parameters p that produce the data d_{obs}.

The inverse problem is unstable $\Rightarrow$ the direct problem is ill-posed

Inverse Problem in the Form of a Fredholm Integral Equation

$$d(x) = \int_{\Omega} K(x,t)p(t)dt,$$

where function p has to be found given the continuous kernel K, function d, and domain Ω in $\mathbb{R}^n$.

- For sufficiently smooth kernels K, the operator F is compact in a reasonable space;
- any solution p (defined up to an additive function lying in the null space) is unstable;
- the Tikhonov regularization is applicable if the solution (not known a priori) has a sufficiently small L^2 norm.

Data-Driven Modeling on Manifolds

- Consider data lying on a low dimensional **manifold** embedded in a high-dimensional Euclidean space $\mathbb{R}^\ell$;
- Show that a space with a **fuzzy partition** has a manifold structure;
- Find the closest manifold suitable for data representation (data-driven aspect);
- Use theory of **F-transforms** as a source of non-local operators including the Laplacian
- Use **non-local Laplacian** in the inverse problem where the corresponding direct is connected with the dimensionality reduction.

Outline

Topological Manifold

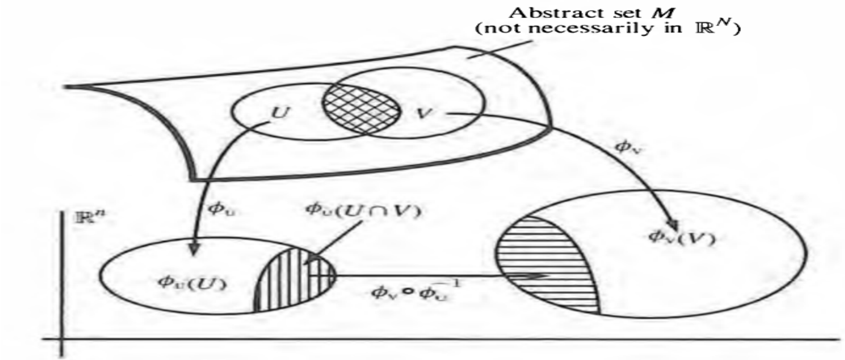
Definition. Gauss, Riemann, Poincare

An n-dimensional topological manifold[2] is a **topological space** M that

- can be covered by a collection of open subsets $\{U_i\}$ which are called **local coordinate neighborhoods** with
- bi-continuous, one-to-one mappings (homeomorphisms) $\phi_i : U_i \to \mathbb{R}^n$, which are called **coordinate maps** (or charts).
- A collection of charts which covers manifold M is called an **atlas** of M. Since all subsets U_i's cover M, we write $M = \bigcup U_i$.
- Since ϕ_i is invertible, the ϕ_i^{-1} exists and it is continuous as well.

[2] J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*, Springer, 2018

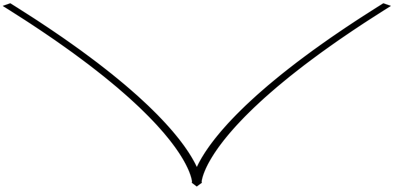
Illustration. Charts



1-Manifold Example

A Cusp - upside down graph of a membership function

The graph of $y = x^{2/3}$ in $\mathbb{R}^2$ is a topological manifold. It is locally Euclidean, because it is homeomorphic to $\mathbb{R}$ via $(x, x^{2/3}) \to x$.



Connected 1-manifolds

Examples of connected 1-manifolds

- The real line $\mathbb{R}$
- The half-line $\mathbb{R}_+$
- The circle $S^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$
- The closed interval $I = [0, 1]$

Topological classification of connected 1-manifolds

Theorem.

Any connected 1-manifold is homeomorphic to one of the four manifolds: $\mathbb{R}$, $\mathbb{R}_+$, S^1, I.

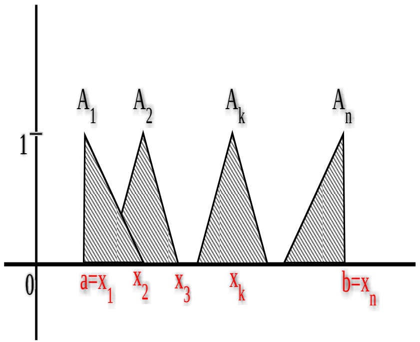
No two of these manifolds are homeomorphic to each other

Outline

Fuzzy Partition $A_1, \ldots, A_n$ of $[a, b]$

Fuzzy sets $A_1, \ldots, A_n$ with **continuous** membership functions form a **fuzzy partition** with nodes $x_1, \ldots, x_n$ if for each $k = 1 \ldots, n$

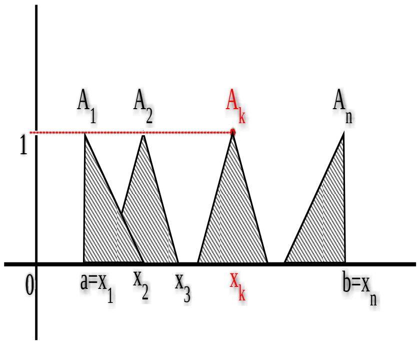
- $A_k(x_k) = 1$
- $A_k(x) = 0$ if $x \notin (x_{k-1}, x_{k+1})$
- $A_k(x) \nearrow$ on $[x_{k-1}, x_k]$
- $A_k(x) \searrow$ on $[x_k, x_{k+1}]$
- Opt-ly, Ruspini condition $\sum_{k=1}^{n} A_k(x) = 1$,



Fuzzy Partition $A_1, \ldots, A_n$ of $[a, b]$

Fuzzy sets $A_1, \ldots, A_n$ with **continuous** membership functions form a **fuzzy partition** with nodes $x_1, \ldots, x_n$ if for each $k = 1 \ldots, n$

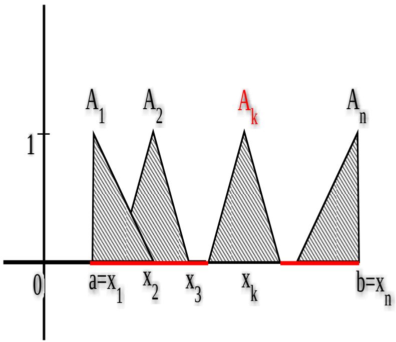
- $A_k(x_k) = 1$
- $A_k(x) = 0$ if $x \notin (x_{k-1}, x_{k+1})$
- $A_k(x) \nearrow$ on $[x_{k-1}, x_k]$
- $A_k(x) \searrow$ on $[x_k, x_{k+1}]$
- Opt-ly, Ruspini condition $\sum_{k=1}^{n} A_k(x) = 1$,



Fuzzy Partition $A_1, \ldots, A_n$ of $[a, b]$

Fuzzy sets $A_1, \ldots, A_n$ with **continuous** membership functions form a **fuzzy partition** with nodes $x_1, \ldots, x_n$ if for each $k = 1 \ldots, n$

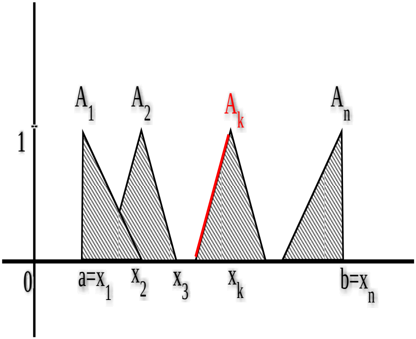
- $A_k(x_k) = 1$
- $A_k(x) = 0$ if $x \notin (x_{k-1}, x_{k+1})$
- $A_k(x) \nearrow$ on $[x_{k-1}, x_k]$
- $A_k(x) \searrow$ on $[x_k, x_{k+1}]$
- Opt-ly, Ruspini condition $\sum_{k=1}^{n} A_k(x) = 1$,



Fuzzy Partition $A_1, \ldots, A_n$ of $[a, b]$

Fuzzy sets $A_1, \ldots, A_n$ with **continuous** membership functions form a **fuzzy partition** with nodes $x_1, \ldots, x_n$ if for each $k = 1 \ldots, n$

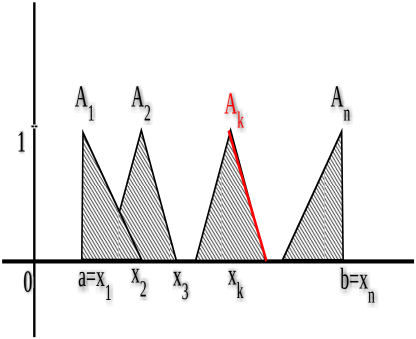
- $A_k(x_k) = 1$
- $A_k(x) = 0$ if $x \notin (x_{k-1}, x_{k+1})$
- $A_k(x) \nearrow$ on $[x_{k-1}, x_k]$
- $A_k(x) \searrow$ on $[x_k, x_{k+1}]$
- Opt-ly, Ruspini condition $\sum_{k=1}^{n} A_k(x) = 1$,



Fuzzy Partition $A_1, \ldots, A_n$ of $[a, b]$

Fuzzy sets $A_1, \ldots, A_n$ with **continuous** membership functions form a **fuzzy partition** with nodes $x_1, \ldots, x_n$ if for each $k = 1 \ldots, n$

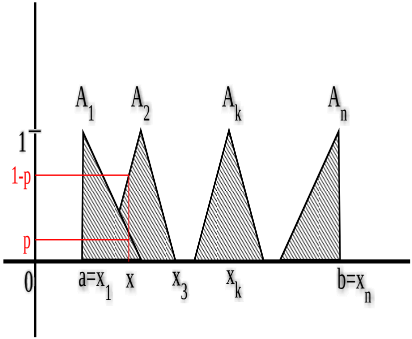
- $A_k(x_k) = 1$
- $A_k(x) = 0$ if $x \notin (x_{k-1}, x_{k+1})$
- $A_k(x) \nearrow$ on $[x_{k-1}, x_k]$
- $A_k(x) \searrow$ on $[x_k, x_{k+1}]$
- Opt-ly, Ruspini condition $\sum_{k=1}^{n} A_k(x) = 1$,



Fuzzy Partition $A_1, \ldots, A_n$ of $[a, b]$

Fuzzy sets $A_1, \ldots, A_n$ with **continuous** membership functions form a **fuzzy partition** with nodes $x_1, \ldots, x_n$ if for each $k = 1 \ldots, n$

- $A_k(x_k) = 1$
- $A_k(x) = 0$ if $x \notin (x_{k-1}, x_{k+1})$
- $A_k(x) \nearrow$ on $[x_{k-1}, x_k]$
- $A_k(x) \searrow$ on $[x_k, x_{k+1}]$
- Opt-ly, **Ruspini condition** $\sum_{k=1}^{n} A_k(x) = 1$,



Fuzzy Partition $\Rightarrow$ 1-Manifold

Let fuzzy sets $A_1, \dots, A_n$ be a non-overlapping **fuzzy partition** of $[a, b]$ with nodes $x_1, \dots, x_n$ where $A_k : (x_{k-1}, x_{k+1}) \to [0, 1]$. Then the collection of charts

$$\{(U_1, \phi_1), (U_2, \phi_2), \dots, (U_n, \phi_n)\}$$

where $U_k = \{(x, A_k(x)) \mid x \in (x_{k-1}, x_{k+1})\}$ and $\phi_k : (x, A_k(x) \to x$ is a 1-dimensional manifold $M_{A_1,\dots,A_n}$ with boundaries $(x_1, 0), \dots, (x_n, 0)$ $(a = x_1,\ b = x_n)$, i.e.

$$M_{A_1,\dots,A_n} = \{(U_1, \phi_1), (U_2, \phi_2), \dots, (U_n, \phi_n), (x_1, 0), \dots, (x_n, 0)\}.$$

1-Manifold $\Rightarrow$ Fuzzy Partition

- Let M be a connected 1-dimensional manifold with n boundary points $p_1, \ldots, p_n$, i.e.

$$M = \{(U_1, \phi_1), (U_2, \phi_2), \ldots, (U_n, \phi_n), p_1, \ldots, p_n\},$$

so that

$$\lim_{p \to p_1} \phi_1(p) = a, \quad \lim_{p \to p_n} \phi_n(p) = b, \ a < b.$$

- Let $A_1, \ldots, A_n$ be a fuzzy partition of $[a, b]$ with nodes $x_1, \ldots, x_n$.
- Let $M_{A_1,\ldots,A_n}$ be the corresponding manifold.

Then manifolds M and $M_{A_1,\ldots,A_n}$ are **homeomorphic**.

Outline

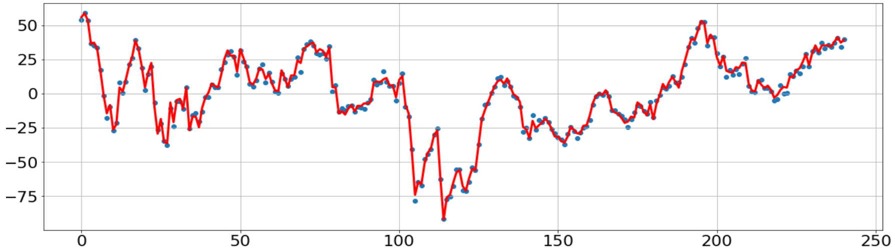
Motivating Problem

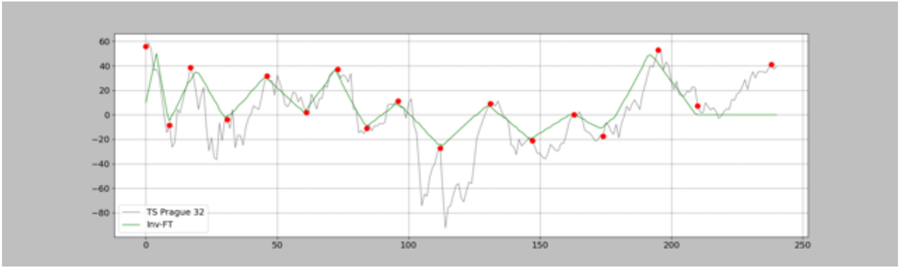
- A continuous function of p, $p \geq 1$, variables on a bounded domain can be represented by its projections (components of the F-transform) onto the nearest manifold with a finite number of connected components, so that the charts of the latter can be used to approximate the function by its inverse F-transform.
- The goal is:

To find the nearest manifold and use it in the invFT

HOW THIS CAN BE DONE?

Motivating Example





Outline

Direct F-transform - details

Definition

Assume that

- $x \in L_2(\mathbb{R})$,
- $\{A_k,\, k \in \mathbb{Z}\}$ is an h-uniform fuzzy partition of $\mathbb{R}$.

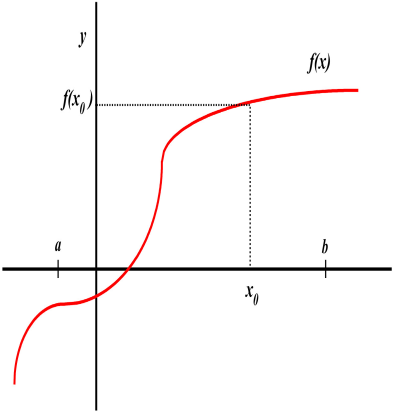
The sequence $F[x] = (X_k,\, k \in \mathbb{Z})$, where

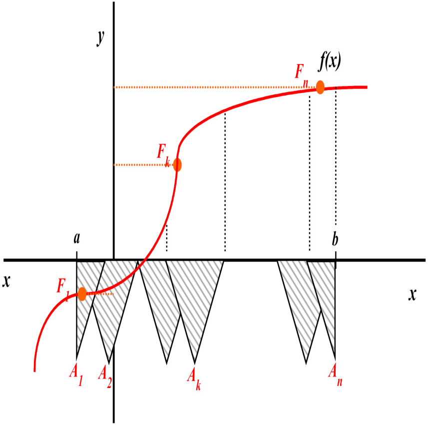
$$X_k = \frac{\int_{-\infty}^{\infty} A_k(s) \cdot x(s)\, ds}{\int_{-\infty}^{\infty} A_k(s)\, ds} = \frac{1}{H} \int_{-\infty}^{\infty} A_k(s) \cdot x(s)\, ds$$

is the *(direct) F-transform* of x with respect to $\{A_k,\, k \in \mathbb{Z}\}$[3]. Real numbers $X_k, k \in Z$, are the *F-transform components* of x.

[3] I. Perfilieva, Fuzzy transforms: Theory and applications Fuzzy Sets and Systems 157 (2006) 993 – 1023

(Fuzzy) F-Transform Schematically





Main Properties of the F-Transform

Best Approximation

Component F_k, $k = 1, \ldots, n$, **minimizes the following criterion**

$$\Phi_k(y) = \int_{-\infty}^{\infty} (f(x) - y)^2 A_k(x) dx.$$

Main Properties of the F-Transform

F-Transform of Constants

Components C_k, $k = 1, \ldots, n$, of a **constant function** c coincide with c, i.e.

$$\mathbf{F}_n(c) = (c, \ldots, c).$$

Main Properties of the F-Transform

Linearity

The F-transform is an image of a linear operator $\mathbf{F}_n$, i.e. for all $f, h \in (L_2[a,b]; A_1, \ldots, A_n)$ and for all $\alpha, \beta \in \mathbb{R}$,

$$\mathbf{F}_n(\alpha f + \beta h) = \alpha \mathbf{F}_n(f) + \beta \mathbf{F}_n(h).$$

Inverse F-transform - details

Inversion Formula

Let

- $\mathbf{x} = (X_k,\, k \in \mathbb{Z})$ be an arbitrary sequence of reals,
- $\{A_k,\, k \in \mathbb{Z}\}$ be an h-uniform fuzzy partition of $\mathbb{R}$,

and The following *inversion formula*[4]

$$\hat{\mathbf{x}}^F(t) = \frac{\sum_{k=-\infty}^{\infty} X_k \cdot A_k(t)}{\sum_{k=-\infty}^{\infty} A_k(t)},\, t \in \mathbb{R},$$

converts the sequence $\mathbf{x}$ into the real function $\hat{\mathbf{x}}^F$ such that $\hat{\mathbf{x}}^F : \mathbb{R} \to \mathbb{R}$.

Definition

$\hat{\mathbf{x}}^F$ is the *inverse F-transform of the sequence* $\mathbf{x} = (X_k,\, k \in \mathbb{Z})$ with respect to the fuzzy partition $\{A_k,\, k \in \mathbb{Z}\}$.

[4] I. Perfilieva, M. Holcapek, V. Kreinovich, A new reconstruction from the F-transform components, FSS 2016

The nearest 1-manifold. Specification

Parameters To be Found

- Boundary points - Keypoints;
- Collection of charts $\{(U_1, \phi_1), (U_2, \phi_2), \ldots, (U_n, \phi_n)\}$ - Local pieces of topologically close points;
- Measure of "goodness".

This Requires to Establish a Calculus on a Manifold !

Weighted Graph as a Model of a Manifold

- Let $G = (V, E, w)$ be a a weighted graph where $V = \{v_1, \ldots, v_\ell\}$ is a finite set of vertices, and E ($E \subset V \times V$) is a set of weighted edges so that $w : E \to \mathbb{R}_+$.
- The edge $e = (v_i, v_j)$ connects two vertices v_i and v_j, and then the weight of e is $w(v_i, v_j)$ or just w_{ij}.
- Weights are set using the function $w : V \times V \to \mathbb{R}_+$, which is symmetric ($w_{ij} = w_{ji}, \forall\, 1 \leq i, j \leq \ell$), non-negative ($w_{ij} \geq 0$) and $w_{ij} = 0$ if $(v_i, v_j) \notin E$.
- The notation $v_i \sim v_j$ denotes two adjacent vertices v_i and v_j with an existing edge connecting them.

Two Hilbert spaces on a Weighted Graph

- Let $H(V)$ denote the Hilbert space of real-valued functions on the set of vertices V of the graph, where if $f, h \in H(V)$ and $f, h : V \to \mathbb{R}$. The **inner product**

$$\langle f, h \rangle_{H(V)} = \sum_{v \in V} f(v)h(v).$$

- Similarly, $H(E)$ denotes the space of real-valued functions defined on the set E of edges of a graph G. This space has the **inner product**

$$\langle F, H \rangle_{H(E)} = \sum_{(u,v) \in E} F(u,v)H(u,v) = \sum_{u \in V} \sum_{v \sim u} F(u,v)H(u,v),$$

where $F, H : E \to \mathbb{R}$ are two functions on $H(E)$.

Outline

Directional Derivative of a Function on a Graph

Let $G = (V, E, w)$ be a weighted graph, and let $f : V \to \mathbb{R}$ be a function in $H(V)$. The difference operator $d : H(V) \to H(E)$ of f, is defined on $(u, v) \in E$ by

$$(df)(u,v) = \sqrt{w(u,v)}\,(f(v) - f(u))\,.$$

The **directional derivative** of f, at vertex $v \in V$, along the edge $e = (u, v)$, is defined as:

$$\partial_v f(u) = (df)(u,v).$$

Adjoint to the Difference Operator

The adjoint to the difference operator $d^* : H(E) \to H(V)$, is a linear operator defined by:

$$\langle df, H \rangle_{H(E)} = \langle f, d^* H \rangle_{H(V)},$$

for any function $H \in H(E)$ and function $f \in H(V)$.

Proposition

The adjoint operator d^* can be expressed at a vertex $u \in V$ by the following formula:

$$(d^* H)(u) = \sum_{v \sim u} \sqrt{w(u,v)}\, (H(v,u) - H(u,v)) \,.$$

The divergence operator, defined by $-d^*$, measures the network outflow of a function in $H(E)$, at each vertex of the graph.

The Weighted Laplace Operator

The weighted gradient operator of $f \in H(V)$, at vertex $u \in V$, $\forall (u, v_i) \in E$, is a column vector:

$$\nabla_w f(u) = (\partial_v f(u) : v \sim u)^T = (\partial_{v_1} f(u), \ldots, \partial_{v_k} f(u))^T.$$

The weighted Laplace operator $\Delta_w : H(V) \to H(V)$, is defined by:

$$\Delta_w f = -\frac{1}{2} d^*(df).$$

w-Laplace Operator and Measure of "Goodness"

Proposition

The weighted Laplace operator Δ_w at $f \in H(V)$ acts as follows:

$$(\Delta_w f)(u) = -\sum_{v \sim u} w(u,v)(f(v) - f(u)).$$

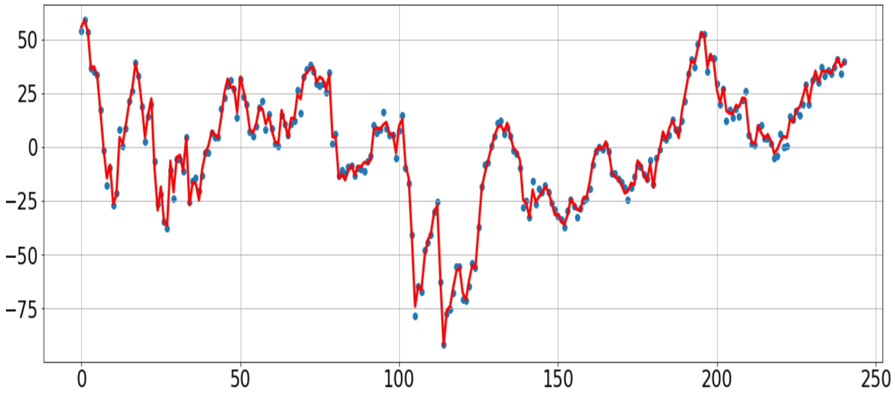
This Laplace operator is linear and corresponds to the graph Laplacian.

Keypoints are the points of local extrema of a Laplacian !

Laplacian values regulate the charts areas !

Outline

Financial Time Series

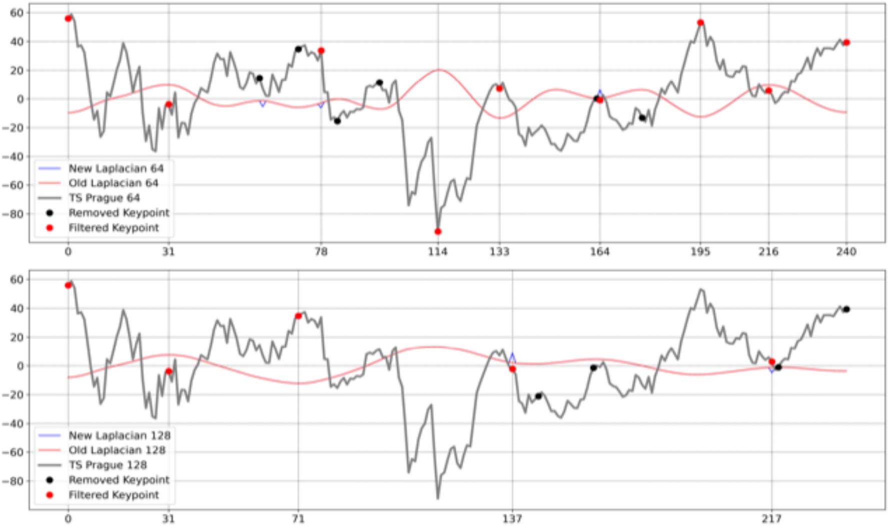


5

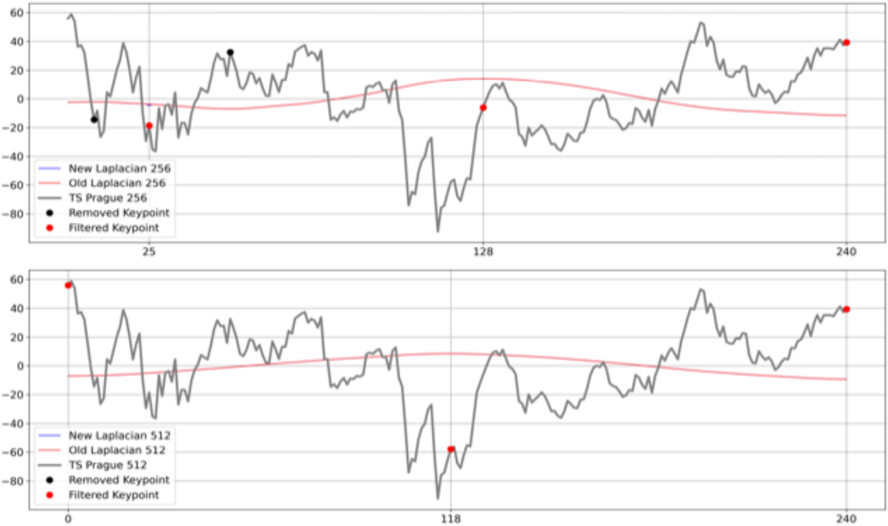
[5] Yahoo Finance: The (2016) daily closing prices from international stock indices, namely Prague (PX), Paris (FCHI), Frankfurt (GDAXI) and Moscow (MOEX)

Scale-Dependent Keypoints, $w = 2^t$, $t = 4, 5$

Scale-Dependent Keypoints, $w = 2^t$, $t = 6, 7$



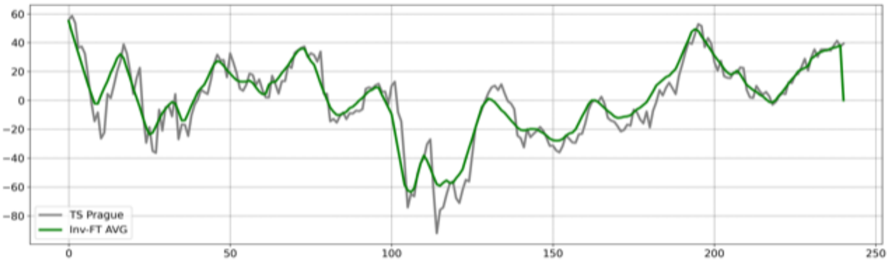
Scale-Dependent Keypoints, $w = 2^t$, $t = 8, 9$

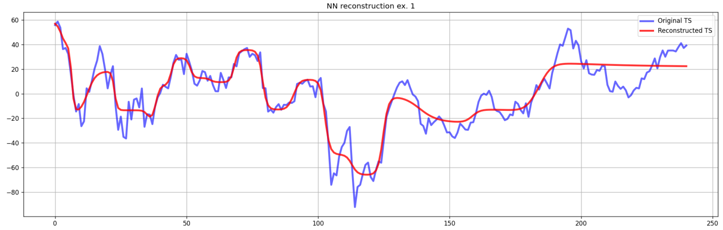


The nearest 1-manifold, $w = 2^t$, $t = 4, 5$

The nearest 1-manifold, $w = 2^t$, $t = 6, 7, 9$

Aggregated Reconstructions: AggIFT – NN





Aggregated Reconstructions with RMSE

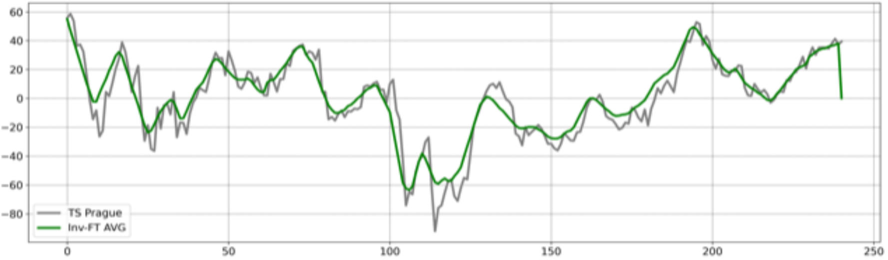


Image and Its Scale-Dependent Keypoints

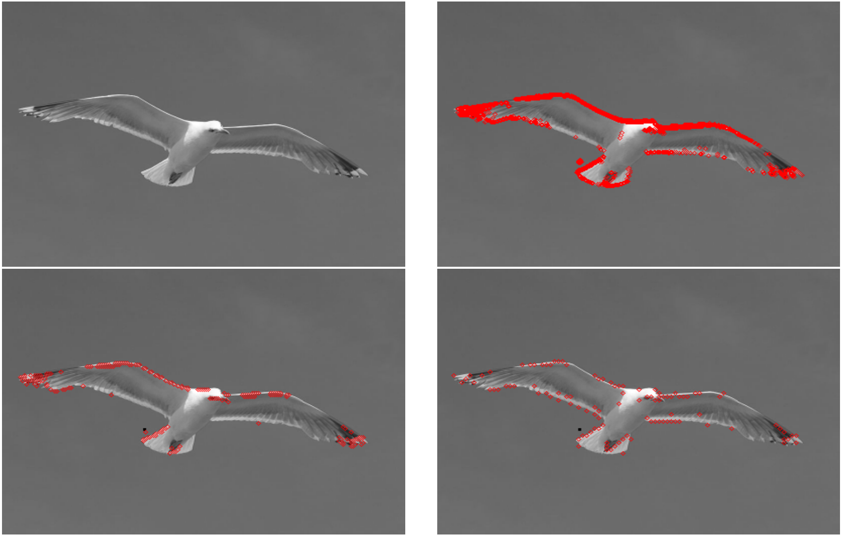
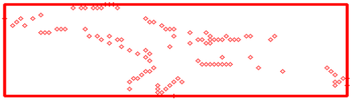


Image and Its Reconstruction from Scale-Dependent Keypoints



Conclusion

- We have contributed to **efficient data-driven modeling** by showing that
  - A connected 1 manifold naturally leads to a space with a fuzzy partition;
  - The data-driven modeling is about finding the nearest manifold;
  - The quality of a data-driven modeling is connected with the Laplace-Beltrami operator.
  - A continuous function on a bounded domain can be represented by its projections (components of the F-transform) onto the nearest manifold, so that the charts of the latter can be used to approximate the function using the inverse F-transform.