Application of Graph Learning to inverse problems

Master Thesis Preperation

Natural Science Faculty of the University of Basel
Department of Mathematics and Computer Science
Data-Analytics
Webpage

Examiner: Prof. Dr. Ivan Dokmanić
Supervisor: Dr. Valentin Debarnot

Cédric Mendelin
cedric.mendelin@stud.unibas.ch
2014-469-274

29.11.2021

Table of Contents

1

Papers - Foundation

1.1 Maths Foundation

1.1.1 Hilbers Space

1.1.2 SO(3), S,

1.1.3 LIE Group?

1.1.4 Principle component analysis - PCA

1.1.5 Local PCA

1.1.6 K-means

1.1.7 K-nearest neughboors

1.2 Graph Foundation

1.2.1 Curvature

1.2.2 Manifold Assumption

1.2.3 Point Cloud

1.2.4 Laplace

1.2.5 Pooling

1.2.6 Link prediction

1.2.7 Hyper Graphs

1.2.8 Power Iterations

1.2.9 Manifold Learning

1.2.10 Signal Processing

1.2.11 Nonlinear dimensionality reduction

1.3 Diffusion Maps:

Coifman and Lafon [1] [1]

Dimensionality reduction: In essence, the goal is to change the representation of data sets, originally in a form involving a large number of variables, into a low-dimensional description using only a small number of free parameters.

meaningful structures in data sets: Analogous to the problem of dimensionality reduction is that of finding meaningful structures in data sets. The idea here is slightly different and the goal is to extract relevant features out of the data in order to gain insight and understanding of the phenomenon that generated the data.
Markov Chain:
Random walk:
PageRank: Stationary distribution of random walk
Kernel eigenmap methods: - local linear embedding - Laplacian eigenmaps, - hessian eigenmaps - local tangent space alignement
The remarkable idea emerging from these papers is that eigenvectors of Markov matrices can be thought of as coordinates on the data set. Therefore, the data, originally modeled as a graph, can be represented (embedded) as a cloud of points in a Euclidean space. two major advantages over classical dimensionality redution (PCA, MDS): The first aspect is essential as most of the time, in their original form, the data points do not lie on linear manifolds. The second point is the expression of the fact that in many applications, distances of points that are far apart are meaningless, and therefore need not be preserved.
Unnormalized Graph Laplacian: $L = D - W$
Normalized Graph Laplacian construction: $L_{sym} = D^{-1/2}LD^{-1/2} = I - D^{-1/2}WD^{-1/2}$ $L_{rw} = D^{-1}L = I - D^{-1}W$
Markov chain has a stationary distribution. If graph is connected, stationary is unique. If X is finite, chian is ergodic.
Diffusion distance: Diffusion map ψ embeds the data into the Euclidean space so that in this space, the Euclidean distance is equal to the diffusion distance.
Laplace–Beltrami operator on manifolds
What are diffusion maps

1.3.1 Vector Diffusion Maps (VDM)

[3] VDMis a mathematical and algorithmic generalization of diffusion maps and other non-linear dimensionality reduction methods, such as LLE, ISOMAP, and Laplacian eigenmaps. While existing methods are either directly or indirectly related to the heat kernel for functions over the data, VDM is based on the heat kernel for vector fields.
Main concept: Edge consists of weight and linear orthogonal transformation. If linear orthogonal transformation is big, nodes are more like to be equal. If small, there are different Diffusion is calculated on vectors fields, where tangets are mapped to the manifold. A way to globally connect Local PCAs.
SNR: signal-to-noise-ration
LLE: ISOMAP: Laplacian eigenmaps:

1.3.2 Riemannian Manifold Assumption:

One of the main objectives in the analysis of a high-dimensional large data set is to learn its geometric and topological structure. Even though the data itself is parametrized as a point cloud in a high-dimensional ambient space R^p, the correlation between parameters often suggests the popular "manifold assumption" that the data points are distributed on (or near) a single low-dimensional Riemannian manifold Md embedded in Rp, where d is the dimension of the manifold and $d << p$.

1.3.3 Multi-Frequency Vector Diffusion Maps (MFVDM)

[2] **For a direct link between manifold embedding and tomography, very close to what Ivan explained this morning. If we have a graph denoising method, we will need to compare with this approach (or the original vector diffusion maps).**
Basically same as VDM, but with multiple frequencies per edge.
Diffusion maps (DM) only consider scalar weights over the edges and the vector diffusion maps (VDM) only take into account consistencies of the transformations along connected edges using only one representation of $SO(2)$, i.e. $e^{ia_{i,j}}$. In this paper, we generalize VDM and use not only one irreducible representation, i.e. $k = 1$, but also higher order k up to k_{max}.

1.4 Graph Laplacian Tomography From Unknown Random Projections

A reference that I already mentioned in the first mail: standard approach that we need to compare with. Maybe their setting (2D tomography with unknown angle) is a good setting to start with.

1.5 Non-local means

Image denoising

1.6 Point Clouds

1.6.1 Dynamic graph Cnn for learning on point clouds

One of the few reference related to graph neural network and learning of graph structure.

1.6.2 CryoEm and related

2. Estimation of Orientation and Camera Parameters from Cryo-Electron Microscopy Images with Variational Autoencoders and Generative Adversarial:
learning framework where the manifold embedding is estimated.
3. Computational Methods for Single-Particle Cryo-EM: review around cryo-EM.
This reference doesn't talk about manifold embedding, but it is a nice one if you want to know more about the acquisition system and standard approaches to solve the cryo-EM problem.
3.bis) Single-Particle Cryo-Electron Microscopy: another review similar to the previous one. The section "Mathematical frameworks for cryo-EM data analysis" and especially the subsection MRA (multireference alignement) introduce a toy model that is related to cryo-EM and where the symmetries are of importance.

4. Bispectrum Inversion with Application to Multireference Alignment: for a paper that introduce several algorithms to solve MRA.

2
Introduction

This is the introduction to the thesis template. The goal is to give students a starting point on how to format and style their Bachelor or Master thesis[1].
Please make sure to always use the most current version of this template, by downloading it always from the original git repository:

http://www.github.com/ivangiangreco/unibas-latex

We will use throughout this tutorial some references to Turing's imitation game [5] and the Turing machine [4]. You may be interested in reading these papers.

The package comes with an option regarding the bibliography style. You can include the package with

```
\usepackage[citeauthor]{basilea}
```

to be able to cite authors directly with

```
\citet{turing:1950}
```

If the option is enabled, then the following reference should print Turing [2]: Turing [5]

[1] This document also shows how to use the template.

3

Body of the Thesis

This is the body of the thesis.

3.1 Structure

3.1.1 Sub-Section

3.1.1.1 Sub-Sub-Section

Paragraph

Even Sub-Paragraph This is the body text. Make sure that when you reference anything you use labels and references. When you refer to anything, you normally capitalise the type of object you reference to, e.g. Section 3.1 instead of section 3.1. You may also just use the `cref` command and it will generate the label, e.g., for Section 3.1, we did not specify the word "Section".
Hint: Try to structure your labels as it is done with `sec:my-label` and `fig:machine`, etc.

3.2 Equations

A Turing Machine is a 7-Tuple:

$$M = \langle Q, \Gamma, b, \Sigma, \delta, q_0, F \rangle \tag{3.1}$$

A Turing Machine is a 7-Tuple even if defined in the text, as in $M = \langle Q, \Gamma, b, \Sigma, \delta, q_0, F \rangle$.

3.3 Tables

Some tables can also be used as shown in Table 3.1[2]. Remember that tables might be positioned elsewhere in the document. You can force positioning by putting a `ht!` in the definition.

[2] Table captions are normally above the table.

Table 3.1: Frequency of Paper Citations. By the way: Make sure to put the label always after the caption, otherwise LaTeX might reference wrongly!

Title	f	Comments
The chemical basis of morphogenesis	7327	
On computable numbers, with an application to the ...	6347	Turing Machine
Computing machinery and intelligence	6130	

3.4 Figures

Figures are nice to show concepts visually. For organising well your thesis, put all figures in the Figures folder. Figure 3.1 shows how to insert an image into your document. Figure 3.2 references a figure with multiple sub-figures, whereas the sub-figures are referenced by Fig. 3.2(a), etc.

Missing: Description figure.

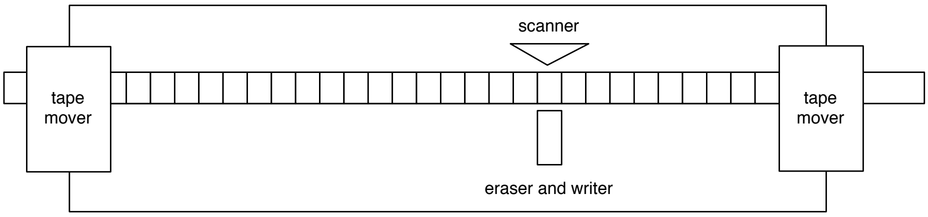
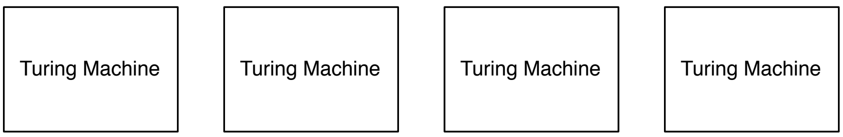


Figure 3.1: A Turing machine.



(a) Turing Machine 1 (b) Turing Machine 2 (c) Turing Machine 3 (d) Turing Machine 4

Figure 3.2: Plots of four Turing machines

3.5 Packages

These packages might be helpful for writing your thesis:

caption to adjust the look of your captions

glossaries for creating glossaries (also list of symbols)

makeidx for indexes and the back of your document

algorithm, algorithmicx, algpseudocode for adding algorithms to your document

4 Conclusion

This is a short conclusion on the thesis template documentation. If you have any comments or suggestions for improving the template, if you find any bugs or problems, please contact me.

Good luck with your thesis!

Bibliography

[1] Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.

[2] Yifeng Fan and Zhizhen Zhao. Multi-frequency vector diffusion maps. In *International Conference on Machine Learning*, pages 1843–1852. PMLR, 2019.

[3] Amit Singer and H-T Wu. Vector diffusion maps and the connection laplacian. *Communications on pure and applied mathematics*, 65(8):1067–1144, 2012.

[4] Alan M Turing. On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London mathematical society*, 42(2):230–265, 1936.

[5] Alan M Turing. Computing machinery and intelligence. *Mind*, 59(236):433–460, 1950.