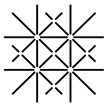


University of Basel

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

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 Introduction

Inverse problems aims at estimating a signal that went through a system, based on the output observation. Machine learning (ML) is a tool to model and solve IP. They are widely used throughout different science directions, such as ML, signal processing, computer vision, natural language processing and many more.
In recent years, Graphs got a lot of attention in ML and are one of the most promising research areas. Graphs are a well suited data structure, simple but with high expressiveness. For some specific scenarios ordinary ML algorithm fail but Graph ML approaches have great success, e.g dimensionality reduction for high-dimensional data. Data can be in a graph structure already, like social networks, or they can be constructed for arbitrary datasets.
Cryo-electron microscopy (cryo-EM), where molecules are imaged in an electron microscope, gained a lot of attention in recent years. Due to ground-breaking improvements regarding hardware and data processing, the field of research has highly improved. In 2017, the pioneers in the field of cryo-EM got the Nobel Prize in Chemistry[1]. Today, using cryo-EM many molecular structures can be displayed with near-atomic resolution. The big challenge with cryo-EM is enormous noise, which makes calculation challenging. During the Master Thesis, the aim is to exploit Graph Learning on the cryo-EM reconstruction problem.

The following report resulted as the Master Thesis Preparation report. During the six weeks project, the aim was to familiarize with the research area, build up some mathematical foundation needed for the Thesis and define the project content as well as a project plan.

The report is structured the following: In chapter **??**, the overall foundation for the Master Thesis will be given, focusing on Graph Learning, Graph Denoising, some mathematical methods and definitions as well as an introduction to cryo-EM. Chapter **??** is dedicated to the problem setup and some preliminaries of the problem. Moreover, the base idea of the Master Thesis is defined. Up to this point, the underlying problem has been defined and some related work can be given in chapter **??**. To end the report, project plan and work packages are introduced in chapter 4.

[1] https://www.nobelprize.org/prizes/chemistry/2017/press-release/

2
Imaging methods

In the current chapter, imaging methods *computed tomography* and *cryo-electron microscopy* (cryo-EM) will be introduced. Further, their observation model is defined in a mathematically way and their reconstruction is observed. Application of cryo-EM is the major motivation for the Master Thesis, as the problem is not easy to solve due to dealing with enormous noise and other difficulties.

2.1 Computed tomography

Computed tomography is a well established imaging method. Using X-ray source, fan shaped beams are produced which scan the imaging object, resulting in many measurements taken over straight lines [3].

Tomography reconstruction: Tomographic reconstruction[6] is a popular inverse problem. The aim is to reconstruct an imaged object from observed measurements. The reconstruction object can be in two-dimension (2D) or in three-dimension (3D). In the Master Thesis, the focus on computed tomography will be on 2D case, which is called *classical tomography reconstruction.*

2D tomographic reconstruction: Mathematically, the observed measurements can be defined as follows:

$$y_i[j] = R(x, \theta_i, s_j) + \eta_i[j], \text{ with } 1 \leq i \leq N, \text{ and } 1 \leq j \leq M \tag{2.1}$$

where N is number of observations and M the observation dimension. Then, $x \in L^2(\Omega)$ is the original object with $\Omega \subset \mathbb{R}^2$ and L^2 is the Lebesgue space. Further, $y_i \in \mathbb{R}^M$ is the i-th observation with $y_i[j] \in \mathbb{R}$ the j-th element of the observation. $R(\cdot; \theta, s) : L^2(\Omega) \to L^2(\tilde{\Omega}), x \mapsto R(x; \theta, s)$ refers to the Radon Transform with $\tilde{\Omega} \subset \mathbb{R}$, θ as the observation angle from the x-axis and s_j as the sampling point. η refers to noise and is defined as $\eta_i[j] \sim \mathcal{N}(0, \sigma^2)$.

Filter Backprojection: Filter Backprojection [6] is a reconstruction method, typically used in classical tomography reconstruction. It allows to inverse the Radon Transform and enables reconstruction of the original object. The algorithm fails when working with noisy data.

2.2 Cryo-EM

Cryo-EM is another imaging method, that enables the view of molecules in near-atomic resolution. In the Master Thesis, only single-particle cryo-EM[8] is considered, when writing from cryo-EM it always refer to single-particle cryo-EM. During imaging process, molecules are frozen in a thin layer of ice, where they are randomly oriented and positioned. Random orientation and positioning of molecules makes reconstruction challenging but the freezing process allows to observe molecules in a stable state where they are not moving. With an electron microscope, two-dimensional tomographic projection images of the molecules are observed in the ice, which are called *micrograph*. The frozen molecules are fragile and the electron microscope needs to work with very low power (electron dose), resulting in highly noisy images. The resulting signal-to-noise ration (SNR) is typically smaller than 1, which indicates that there is more noise than signal[14]. Further, observed molecules are not equal is the sense that there are some structural varieties.

3D cryo-EM reconstruction: Similar to tomographic reconstruction, cryo-EM reconstruction problem[2] is defined. It can be seen as a 3D reconstruction problem as the original object $x \in L^2(\Omega)$ to be reconstructed is in 3D. To keep the notation from previous section, now $\Omega \subset \mathbb{R}^3$ and $\tilde{\Omega} \subset \mathbb{R}^2$.
Mathematically, the observed measurements can be defined as follows:

$$y_i = \Pi_z(Rot(x;\theta_i)) + \eta_i, \text{ with } 1 \leq i \leq N \tag{2.2}$$

where $\Pi : L^2(\Omega) \to L^2(\tilde{\Omega}), x \mapsto \int x(\cdot,\cdot,z)dz$ is the projection operator and $Rot : L^2(\Omega) \to L^2(\Omega), Rot_\theta(x) = \big((x_1,x_2,x_3) \mapsto x(x_1R^1, x_2R^2, x_3R^3)\big)$ is the rotation operator modelling the rotation during freezing. Further, $\theta_i = [\theta_i^1, \theta_i^2, \theta_i^3]$ where entries $[\theta_i^1, \theta_i^2, \theta_i^3 \in \mathbb{R}$ and $R = [R^1, R^2, R^3] \in SO(3)$. $\eta_i[j,k] \sim \mathcal{N}(0, \sigma^2 I)$ corresponds again to the noise of the observation.
As y_i is not observable directly, discretization is needed:

$$\begin{aligned} y_i &= (\Pi_z(Rot(x;\theta_i)) + \eta_i)(\Delta), \text{ with } 1 \leq i \leq N \\ y_i[j,k] &= \Pi_z(Rot(x;\theta_i))_{j,k} + \eta_i[j,k], \text{ with } 1 \leq i \leq N \text{ and } 1 \leq j,k \leq M \end{aligned} \tag{2.3}$$

where $\Delta \subset \tilde{\Omega}^{M^2}$ is the sampling grid and M is the first and second dimension of the sampling grid.

Extended formula: The equation 2.2 is a simplified version of the cryo-EM reconstruction problem. First of all, the point spread function (PSF) of the microscope is not taken into account. Moreover, the structural variety is not taken into account, the underlying object x is not the same for every observation but can be seen as a random signal from an unknown distribution defined over all possible molecules structures.

The extended version can be defined as follows:

$$y_i = h_i \circ \Pi_z(Rot(x_i; \theta_i)) + \eta_i, \text{ with } 1 \leq i \leq N \tag{2.4}$$

where h_i is the PSF of the microscope and $\circ$ defines the convolution. During the Master Thesis, the equation **??** is used, not the extended version.

Difference to tomographic reconstruction: The two problems are highly related, but the cryo-EM reconstruct is more challenging. During CT observation, the patient is asked to not move and therefore, the angles of projection are known. Whereas, in cryo-EM this information will be lost during the freezing process. Secondly, the high level of noise makes cryo-EM much more challenging regarding tomographic reconstruction.

2.3 Abstract from

As the tomographic reconstruction and the cryo-EM reconstruction are rather similar, the aim of the Master Thesis will be to design an algorithm, that can be applied in both scenarios. Therefore, a abstract form of the problems will be defined in the following. First of all, a similar notation as before is used, but in a more general way $x \in L^2(\Omega)$ where $\Omega \subset \mathbb{R}^D$ with D as the dimension of the space and $\tilde{\Omega} \subset \mathbb{R}^{D-1}$.

$$y_i = \left(A(x, \theta_i) + \eta_i\right)(\Delta) \tag{2.5}$$

where $y_i \in \tilde{\Omega}^M$ is the observed measurements, M the measurement dimension, $x \in L^2(\Omega)$ our original object, A a non-linear operator $A : L^2(\Omega) \to L^2(\tilde{\Omega}), x \mapsto A(x; \theta)$ and $\eta\ \mathcal{N}(O, \sigma^2 I)$ gaussian noise. $\Delta \subset \tilde{Omega}^{M^2}$ is a term for discretization.

Classical tomography reconstruction: For classical tomography parameters are defined defined with $D = 2$ and $\theta \in \mathbb{R}^1$. Further, $A(\cdot)$ is the Radon transform, defined in equation 2.1. A distance measure between measurements can be set up by using the l2-norm $\|y_i - y_j\|$.

Cryo-Em reconstruction: For cryo-EM parameters are defined with $D = 3$ and $\theta \in \mathbb{R}^3$. Further, $A(\cdot)$ can be defined as $\Pi_z\left(Rot(x; \theta)\right)$ where Rot is the 3D rotation and Π_z the tomographic projection.
As measurements are drawn with some random 3D rotation and projection, it can happen that two samples are equivalent up to 2D rotation. Consider a first example y_1, which has no 3D rotation and a second sample y_2 with a rotation only in in x-y plane by 45°. The two samples have a defined in-plane rotation g, such that $gy_1 = y_2$. Therefore, in our distance measure we add this term of in-plan rotation: $min_{g \in SO(1)} \|g * y_i - y_j\|$, which is inspired by the work of [9].

High noise regime: Cryo-EM measurements are highly noisy, which makes reconstruction challenging. There are different ways to reduce noise from measurements, most of them are related to averaging. Averaging need to consider similar measurements and ignore

diverse ones. In the defined abstract model, averaging over paired measurements from θ should be a good averaging model. But how can it be achieved?
One idea would be to measure distances between observation (therefore introduced above). Another way is to find a low-dimensional embedding which maps our measurements y to some θ. When talking from low-dimensional embeddings, there is no way around Graph Learning, which will be introduced in the following chapter.

During the Master Thesis, high-noise regime is the domain of interest. The main practical application is cryo-EM, where an algorithm for denoising is expected to boost quality of the overall 3D-reconstruction. The aim of the Master Thesis is to introduce a denoising algorithm, which is able to work well even on highly noisy data, where cryo-EM is major field of interest.

3
Graph Denoising

In the following chapter, the connection between graphs and the domain of denoising in high-noise domains, such as cryo-EM, is established. First, a broad definition of graphs is given and further, the term "Graph Denoising" is introduced and explained. Finally, the connection to Graph Laplacian is established and different opportunities exploiting for a good denoising algorithms are shown.

3.1 Graph Foundations

Graphs are a powerful representation of data, simple but with high expressiveness. Real world data could be in graph structure already, like social networks, citation networks, protein interaction networks or google search. If not already in graph structure, a graph can be artificially constructed with methods like k-nearest neighbours (k-NN) or others.

Graph Learning: is a popular research area and got a lot of attention in recent years. It is a way of applying Machine Learning (ML) on graphs and algorithms emerged from ML but also other fields. When a graph is available, one can start using Graph Learning algorithms for solving tasks. Popular tasks are node classification or link prediction within a graph. One tries to learn from node and edge features as well as the topology of the graph and tries to map information to a model, which allows prediction or classification. Another popular task in Graph Learning is community detection, where the aim is to identify cluster of nodes within the input graph. Further, graphs are highly popular for dimensionality-reduction, where graph algorithms provide a helpful tool, as ordinary algorithms like principle component analysis fail.

3.1.1 Graph definition

A graph is defined as $G = \langle V, E \rangle$, where V is a set of vertices (or nodes) and E is a set of edges (or links). Edges are defined as a set of tuples (i, j), where i and j determine the index of the vertices in the graph.

Graph properties: A graph can be either *directed* or *undirected*. In a directed one, a edge points explicitly from one node to another, which means that edge $(i, j) \neq (j, i)$. In undirected graphs the ordering does not matter and $(i, j) = (j, i)$.
The *neighbourhood* of a node $\mathcal{N}(\rangle) = \forall x$ is defined as all adjacent nodes of i, or in other words, there is an edge between the nodes. Further, edges can have *weights*, which is a method to define importance to neighbours of a node. If edges are dealing with weights, the term *weighted* graph is used. The *degree* of a node are the number of incoming edges.

Adjacency matrix: To do calculations with graphs, it is common to translate graphs in a well suitable mathematically form, which are matrices. The adjacency matrix can be seen as a way of representing graphs as a matrix. The (binary) adjacency matrix of graph G is defined as follows:

$$A_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0, & \text{otherwise} \end{cases} \tag{3.1}$$

The matrix A has dimension $\mathbb{R}^{N \times N}$ and the indices of the matrix correspond to the nodes of the graph. If there is an edge between two nodes, the entry in the matrix will be set to 1, otherwise to 0. This leads to an unweighted graph, as the weight of all edges will be 1, but could easily be extended by assigned not just values of 1 or 0.
When the graph is undirected, the resulting matrix will be symmetric. Eigenvalues of A are also called *spectrum* of the graph.

3.1.2 Graph Construction

3) General framework for constructing a graph: a) Each vertex is associated to some feature/signal/whatever $x \in \mathcal{X}$, for some space arbitrary space $\mathcal{X}$. b) We construct the graph G_0 using: $d(x_i, x_j) < \tau$, τ is a threshold, or k-NN (defined it here in one line). c) For simplification, and because it covers already most cases, we will work on $\S = R^M$ d) There are some applications, as we will see (or have seen if already introduced), where we have only access to a noisy version of the true signal: $y = x + \eta$, where $y, x \in R^N$, and eta i.i.d follow Gaussian $(0, sigma^2)$. e) Then we defined the noisy graph G as in b)
When data is not available as a graph, it can be constructed from the data. Consider data from space $\Omega = \mathbb{R}^M$, but could basically be any arbitrary space. Then, each node is associated with some element $x \in \otimes$. Further, the graph G can be constructed by using:

$$A_{ij} = \begin{cases} 1 & \text{if } d(x_i, x_j) < \tau \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

where n_i, n_j are nodes from indices i, j, d is a similarity measure between two nodes and τ is a threshold, when to consider two nodes to be adjacent. K-NN is one possible implementation, where for every node, k neighbours will be defined The neighbourhood of node i is defined as $\mathcal{N}_i$ and consists of the nodes, with the k smallest similarity measure.

Noise regime In the case of high noise regime, observation of x it not possible. Measurements will give access to $y = x + \eta$ where $y, x \in \Omega$ and the noise η drawn from gaussian

distribution $\mathcal{N} \sim (0, \sigma^2)$. The *noisy graph* G_0 can be constructed as in equation 3.2, but replacing y with x:

$$A_{ij} = \begin{cases} 1 & \text{if } d(y_i, y_j) < \tau \\ 0, & \text{otherwise} \end{cases} \tag{3.3}$$

3.2 Graph Denoising definition

First of all, Graph Denoising is not a common term in literature, but it is rather related to signal or image denoising. Reconstruction of a true signal given noisy observation signal is done via averaging, that can be performed locally, by the calculus of variations or in the frequency domain[4].
In the last section, noisy graph G_0 was introduced and the aim is to denoise this graph, which basically means to estimate the original graph G from a given noisy graph G_0.

> The goal of the Master Thesis is to introduce a method to estimate the original graph G based on an observed noisy graph G_0.

Noisy Graph: For every noisy graph, there exists an original graph $G = \langle V, E \rangle$.
The noisy graph can be further defined as follows:

$$\begin{aligned} G_{noisy} &= \langle V, E_{noisy} \rangle, \\ \text{with } E_{noisy} &= E \setminus E^- \cup E^+, \\ E^- &\subseteq E, \\ E^+ \cap E &= \emptyset \end{aligned} \tag{3.4}$$

The noisy graph consists of the same nodes as the original graph. From the original graphs edges, some are removed (denoted by E^-) and some are added (denoted by E^+).
The adjacency matrix of G_{noisy} is denoted by $\bar{A}_{ij}$. The task of Graph Denoising, can therefore be written as:

$$\bar{A} \xrightarrow[method]{Graph-denoising} \tilde{A} \approx A \tag{3.5}$$

Where $\bar{A}$, $\tilde{A}$, A denotes the adjacency matrix from noisy input graph, denoised graph and original graph respectively.

Connection to link prediction Link prediction is a task in Graph Learning. The idea is to predict existence of a link (edge) between two nodes. The task can be formulated as a missing value estimation task. A model M_p is learned from a given set of observed edges. The model finally maps links to probabilities $M_p : E' \to [0, 1]$ where E' is the set of potential links.
We define U as the set of all possible vertices of G, therefore $E \subseteq U$. Obviously, graph denoising can be seen as a link prediction problem.
The difference is, that in link prediction a model from a set of observed links is learned $E_{observed} \subseteq E$ and in Graph Denoising the model is learned from $E_{observed} \subseteq U$.

3.2.1 Non local means:

Non local means is a state-of-the-art image denoising method [4]. For a given noisy image v, the denoised image is defined as $NL[v](i) = \sum w(i,j)\, v(j)$. where $w(i,j)$ is the weight between pixel i and j. The weight can be seen as a similarity measure of the two pixels. Moreover, these similarities are calculated over square neighbourhoods of two pixels, where the $\ell 2$-norm of the neighbourhood is used. Similar pixel neighbourhoods have a large weight and different neighbourhoods have a small weight. More general, the denoised image pixel i is computed as an weighted average of all pixels in the image, therefore, in a non local way.

3.3 Graph Laplacian

The Graph Laplacian is a matrix that represents the graph and can be used to find many important properties of the graph, a good overview can be found by [15, 18]. It is defined as follows:

$$L = D - A, \tag{3.6}$$

where A is the adjacency matrix and D the degree matrix (diagonal matrix with degree of nodes as entries).

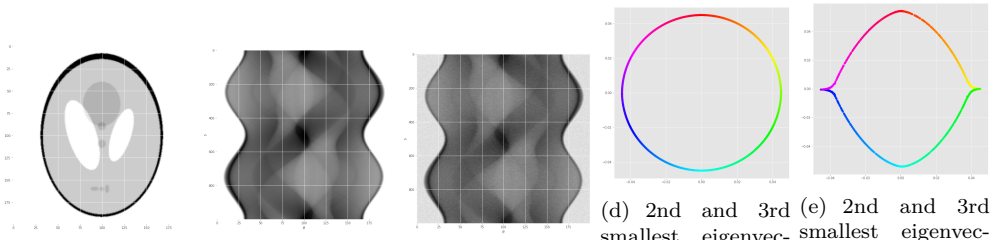
Manifolds: In high-dimensional data euclidean distances are not meaningful in the sense that they will not capture similar data well. Graph Laplacian can be used to compute a *Manifold*, which can help in such a scenario. In the manifold space, euclidean distances make sense again. Let the manifold M be defined as $\mathcal{M} = \{f(x), f \in C^K, f : \mathbb{R}^D \to \mathbb{R}^d\}$. Manifolds are a well established mathematical concept. In the Master Thesis, we will only consider C^k differentiable d-dimensional manifold defined by $\mathcal{M}$. When $d \ll D$, the manifold defines a *low dimensional embedding*, which maps from a high dimensional space $\mathbb{R}^D$ to a low dimensional space $\mathbb{R}^d$.
Lets give two popular examples of such manifolds, namely the *circle* and the *sphere*. The circle is a 1D manifold, where $d = 1$ and $D = 2$. A sphere is a 2D manifold, with $d = 2$ and $D = 3$.

Manifold assumption: The manifold assumption is a popular assumption for high-dimensional datasets. For a given dataset in high-dimension, one can assume that data points are samples drawn from a low-dimensional manifold, that embeds the high-dimensional space. Therefore, if the underlying manifold can be approximated, a dimensionality reduction is established as one can embed the data points in the low-dimensional manifold space. There is a complete area of research devoted to this manifold assumption called Manifold Learning[5].

In the field of classical tomography and cryo-EM, the manifold is well defined for none-noisy data. In the 2D case of classical tomography, the underlying manifold is a circle, whereas in 3D case of cryo-EM the manifold if defined as a sphere. This fact can be exploited during learning, by using the wasserstein loss function (see A.2).

The manifold, and therefore, a low-dimensional embedding, can be calculated the following:

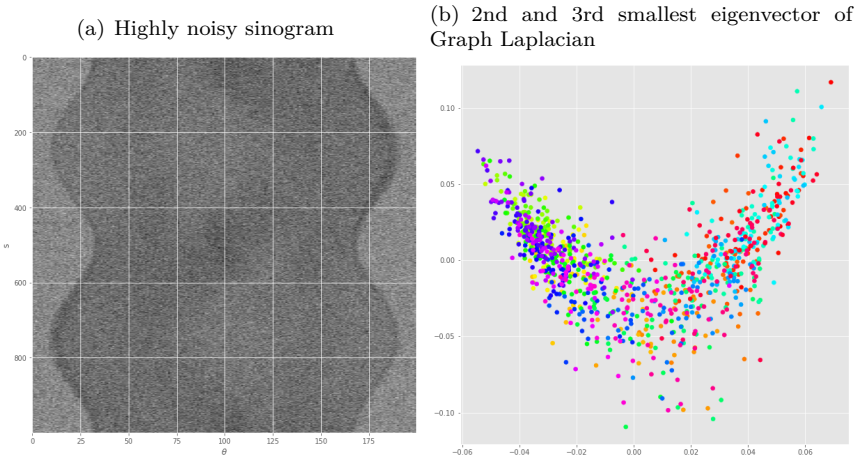
1. Construct the knn-graph from our observations (see section 3.1.2).
2. Calculate the (normalized) Graph Laplacian.
3. Extract the second, third (and fourth) smallest eigenvectors.



(a) Shepp-Logan phantom (b) Original sinogram (c) Noisy sinogram (d) 2nd and 3rd smallest eigenvector of Graph Laplacian (e) 2nd and 3rd smallest eigenvector of noisy Graph Laplacian

Figure 3.1: Shepp-Logan phantom manifold

In Figure 3.1(d) the manifold calculated from the original Graph Laplacian can be seen and it is a perfect circle. Next to it, in Figure 3.1(e) the noisy version with $\sigma = 2$ is plotted and the manifold is circle like but not at all like from the original one.
The more noise we add, the less the manifold looks like a circle. In Figure 3.2(b) the manifold for $\sigma = 100$ is plotted.



(a) Highly noisy sinogram (b) 2nd and 3rd smallest eigenvector of Graph Laplacian

Figure 3.2: Shepp-Logan phantom manifold for high noise level

In all the plots, knn-graph have been constructed with $k = 10$.
TODO: If time, define SNR properly and do some other plots with SNR smaller to 1

3.3.1 Connection to machine learning

Graph Laplacian is used for dimensionality reduction for high-dimensional data, as well as spectral clustering and semi-supervised learning. Coifman et al. [7] used Graph Laplacian in a complete other domain, namely in tomography. They showed that Graph Laplacian approximates the Laplace-Beltrami operator. Further, Graph Laplacian is depended on the adjacency matrix A. If A is noisy, Graph Laplacian will be noisy as well.

In the Master Thesis, new experience of Graph Laplacian in the domain of tomography and denoising should be gained. Further, the connection with machine learning will be explored, hopefully allowing to learn a denoised adjacency matrix to fully enable the power of Graph Laplacian.

3.3.2 Graph Deep Learning

As already mentioned, Graph Denoising can be seen as a way of link predication. The state-of-the-art method for solving link prediction are graph deep learning approaches. Graph deep learning is a fast evolving field in research. With Graph Neural Networks (GNN)[12] the framework for GNN has been established.

Using Graph Convolutional Networks (GCN) [13] for graph feature extraction is a popular way. Basically, with GCN a new feature representation is iteratively learned for the node features (edges features are not taken into account). It can be seen as an averaging of nodes over their neighbourhood where all the neighbours get the same weight combined with some non-linear activation. To consider the node itself in the averaging process they apply to so-called "Renormalization trick", where self-loops are added to the adjacency matrix and after every layer, a normalization step is applied.The topology of the graph will not be adjusted during the learning process.

Veličković et al. [17] extended the concept of GCN with attention and not all the neighbouring nodes get the same weight (attention). Simple Graph Convolutional Network (SGC) [21] proposed a simplified version of GCN. They could verify their hypothesis that GCN is dominated by the local averaging step and the non-linear activation function between layers do not contribute to much to the success of GCN. Therefore, it can be seen as a way of power iteration (see A.1 for further information) over the adjacency matrix with normalization in every layer. Wang et al. [20] proposed a extended version of GCN by not operating on the same graph in every layer but adopting the underlying graph topology layer by layer.

In the Master Thesis, the connection between Graph Laplacian and GNNs should be further exploited. The fact, that [21] could simplify the existing GCN algorithm is motivation enough, that similar connections can be drawn in other fields of Graph Learning.

4

Master Thesis project

In the last chapter of the Thesis Preparation report, the project plan will be introduced as well as a broad overview of different work packages. Further, the project timeline can be seen as a Gantt chart. Probably, there are some parts which will not work out as expected and adjustments are needed throughout the Thesis, the project plan can be seen as a rough guideline.

4.1 Problem conclusion:

Conclusion from red boxes.
During the Master Thesis, the reconstruction problem with unknown angles is considered. Moreover, the observed samples are considered to be noisy. The resulting proposed algorithm should work in the 2D and 3D scenario (classical tomography and cryo-Em).
The main idea is to exploit the fact, that the underlying manifold is known (circle in 2D and sphere in 3D). From our noisy observations, a manifold can be computed and compare it with the original manifold. The comparison between the manifolds enables the possibility of a loss function and learning in general.
It is expected, that the folded spectrum [19] introduced in section **??** can be used to estimate the eigenvalues of the Graph Laplacian. Further, as already mentioned in section A.2, the wasserstein metric is a good choice as a loss function when it comes to dealing with data from a manifold distribution [1], as in our case.
The problem can be seen as Graph Denoising as observations are noisy and therefore, the proposed algorithm will denoise the graph based on the manifold assumption.

Evaluation: During evaluation, 2D and 3D scenario will be considered. A first evaluation will be done on artificial constructed toy-dataset. If time allows, real dataset from classical tomography and/or cryo-EM[2] can be evaluated as well. During evaluation, two baselines are considered which already solved the problem. The first one is a multi-frequency diffusion map approach[9, 10], which aims to denoise cryo-EM images. Secondly, [7] a Graph Lapla-

[2] https://www.ebi.ac.uk/emdb/

cian approach solving classical tomography with random projection angles will be compare against. The evaluation process is a first broad idea. Any adjustments in baseline papers or dataset are possible during the Master Thesis. The baseline papers are further addressed in the related work chapter **??** and detailed work packages are defined in chapter 4

4.2 Work packages

Implement algorithm for 2D case: The first step will be, to familiarize with the problem and implement the algorithm for 2D.

Evaluate 2D case on toy dataset and implement baselines: As a second step, the implemented 2D algorithm will be tested on a toy dataset, where noise is added to the images by hand. As the aim is to work with highly noisy images, the noise level can be selected and increased when working with toy datasets. The evaluation in 2D is crucial and needs to in a satisfying matter. It does not make sense to continue with 3D implement, when the simply 2D case is not handled well enough. Therefore, if the evaluation results are not satisfying, the algorithm needs to be iteratively adjusted, such that the evaluation will be in a good enough quality.

Implement algorithm for 3D case: After successfully evaluating the algorithm in 2D, the aim is to extend the algorithm to work in 3D as well.

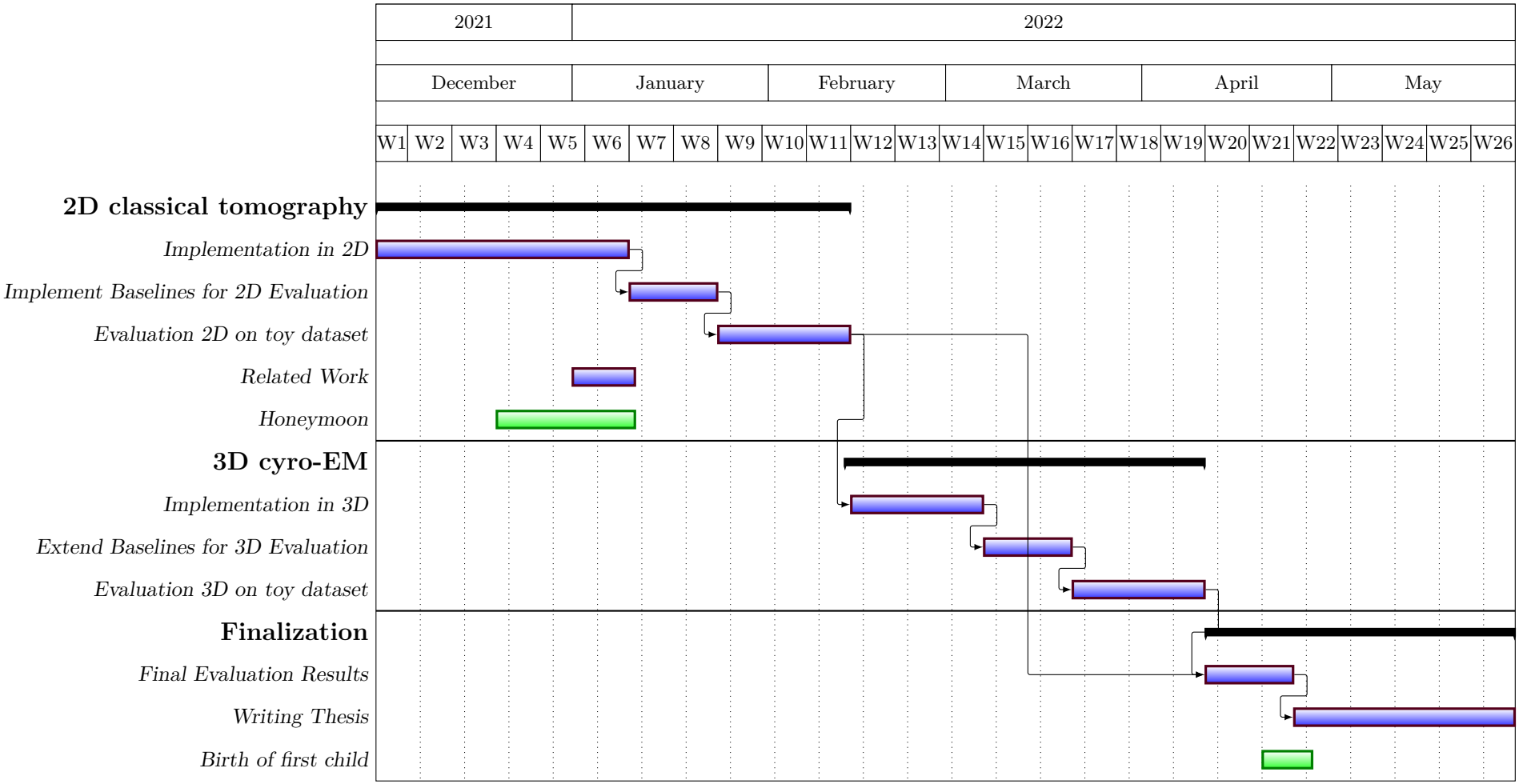
Evaluate 3D case on toy dataset and adjust baselines: Again, the implementation will be evaluated on a toy dataset, where noise can be adjusted by hand.

Nice to have: Evaluate of real dataset If time allows and the 2D and 3D implementation are evaluated successfully on toy datasets, real data can be used for further evaluation. This step will only be done, if time allows.

Evaluate related work: As cryo-EM reconstruction is a hot research topic, related work can not only be considered during the start of the Thesis and needs to be evaluated throughout the Thesis.

Writing Thesis: Document implementation and evaluation result.

4.3 Gantt chart



Bibliography

[1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 214–223. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/arjovsky17a.html.

[2] Tamir Bendory, Alberto Bartesaghi, and Amit Singer. Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities. *IEEE signal processing magazine*, 37(2):58–76, 2020.

[3] David J Brenner and Eric J Hall. Computed tomography—an increasing source of radiation exposure. *New England journal of medicine*, 357(22):2277–2284, 2007.

[4] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65. IEEE, 2005.

[5] Lawrence Cayton. Algorithms for manifold learning. *Univ. of California at San Diego Tech. Rep*, 12(1-17):1, 2005.

[6] Rolf Clackdoyle and Michel Defrise. Tomographic reconstruction in the 21st century. *IEEE Signal Processing Magazine*, 27(4):60–80, 2010.

[7] Ronald R Coifman, Yoel Shkolnisky, Fred J Sigworth, and Amit Singer. Graph laplacian tomography from unknown random projections. *IEEE Transactions on Image Processing*, 17(10):1891–1899, 2008.

[8] Allison Doerr. Single-particle cryo-electron microscopy. *Nature methods*, 13(1):23–23, 2016.

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

[10] Yifeng Fan and Zhizhen Zhao. Cryo-electron microscopy image denoising using multi-frequency vector diffusion maps. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 3463–3467. IEEE, 2021.

[11] Charlie Frogner, Chiyuan Zhang, Hossein Mobahi, Mauricio Araya-Polo, and Tomaso Poggio. Learning with a wasserstein loss. *arXiv preprint arXiv:1506.05439*, 2015.

[12] Marco Gori, Gabriele Monfardini, and Franco Scarselli. A new model for learning in graph domains. In *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, volume 2, pages 729–734. IEEE, 2005.

[13] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[14] Amit Singer. Mathematics for cryo-electron microscopy. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*, pages 3995–4014. World Scientific, 2018.

[15] Daniel Spielman. Spectral graph theory. *Combinatorial scientific computing*, 18, 2012.

[16] Peter Toft. The radon transform. *Theory and Implementation (Ph. D. Dissertation)(Copenhagen: Technical University of Denmark)*, 1996.

[17] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[18] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4): 395–416, 2007.

[19] Lin-Wang Wang and Alex Zunger. Solving schrödinger's equation around a desired energy: Application to silicon quantum dots. *The Journal of Chemical Physics*, 100(3): 2394–2397, 1994.

[20] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*, 38(5):1–12, 2019.

[21] Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Weinberger. Simplifying graph convolutional networks. In *International conference on machine learning*, pages 6861–6871. PMLR, 2019.

A

Mathematical tools

A.1 Power Iterations

Power iteration (also called power method) is an iteratively method, that approximates biggest eigenvalue of a diagonalizable matrix A.
The algorithm starts with a random vector b_0 or an approximation of the dominant eigenvector.

$$b_{k+1} = \frac{Ab_k}{||Ab_k||} \tag{A.1}$$

TODO:convergence if there is only one largest eigenvalue and if b0 is not orthogonal to the eigenvector associated with the largest eigenvalue.
The algorithm not necessarily converges. The algorithm will converge, if A has an eigenvalue strictly grater than its other eigenvalues and the initial vector b_0 has a component in direction of an eigenvector, associated with the dominant eigenvector.

A.2 Wasserstein metric

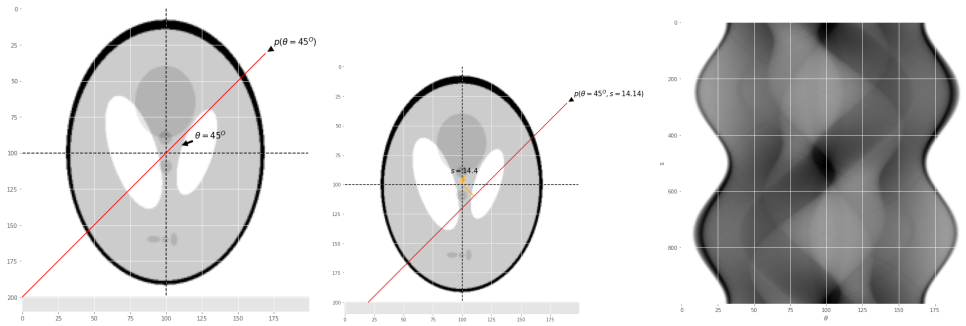
TODO: the difference between Wasserstein and KL divergence for instance is that it is defined (the value is finite) even if the two distributions have not the same support.
The Wasserstein metric is a distance measure between two probability distributions and it is used in ML as a loss function[11]. Intuitively, it can can be understood as the minimum cost to transfer the mass of one distribution to the other. Therefore, it is also known as the *earth mover's distance*.
As Arjovsky et al. [1] could show, ordinary distance measures like *Total Variation*, *Kullback-Leibler divergence* and *Jensen-Shannon divergence* are not sensible when learning with distributions supported by manifolds On the contrary, Wasserstein metric does a good job as loss function in such scenarios.

A.3 Radon Transform

The Radon Transform[16] is the main mathematically concept of tomographic reconstruction. It is an integral transformation and the inverse for classical tomography is well defined by the Fourier Transform.

For classical tomography, $R : f \to Rf, f(x, y) \mapsto Rf(\theta, s)$, where f is a 2D image and x and y can be seen as the coordinates within this image. Then, $Rf(f; \theta, s)$ defines a line, where s is the distance from the origin and θ is the angle to the x-axis.

In Figure A.1(a) and Figure A.1(b) on can see two plots of different values for θ and s, where $f(x, y)$ is the Shepp-Logan phantom. The complete $R(\theta = 45, s = 0)$, which is also called *sinogram*, can be see in Figure A.1(c)



(a) Radon Transform $R(x; \theta = 45, s = 0)$

(b) Radon Transform $R(x; \theta = 45, s = 14.14)$

(c) Shepp–Logan phantom sinogram of $R(x; \theta = 45, s = 0)$

Figure A.1: Examples, where the original object x is the Shepp-Logan phantom.