ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

Scuola di Scienze
Dipartimento di Fisica e Astronomia
Corso di Laurea in Fisica

A MATHEMATICAL INTRODUCTION TO GEOMETRIC DEEP LEARNING

Relatore:
Prof.ssa. Rita Fioresi

Presentata da:
Tommaso Lamma

Anno Accademico 2020/2021

Contents

Chapter 1

Preliminaries on topology

The essential idea in algebraic topology is to convert problems about topological spaces and continuous functions into problems about algebraic objects and their homomorphisms, this way one hopes to end up with an easier problem to solve.

1.1 Simplicial Complexes

In this section we shall define structures called simplicial complexes and discuss some of their properties. In order to define these structures we need the definitions of convex hull and affine independence in $\mathbb{R}^n$. In this chapter we recall some notions of algebraic topology, such as simplicial complexes and homology. For more details we invite the reader to consult [5], a good reference also for the preliminary necessary notions of topology we are unable to treat here.

Definition 1.1.1. Let $A \subset \mathbb{R}^n$, we define A to be *convex* if

$$x, y \in A \Rightarrow tx + (1+t)y \in A$$

for all $t \in [0,1]$.

In Figure 1.1 can see in blue an example of a convex set: every segment joining two points of the set lies within the set. The green set is not convex, in fact we see that the segment in the illustration partially lies outside of the set.

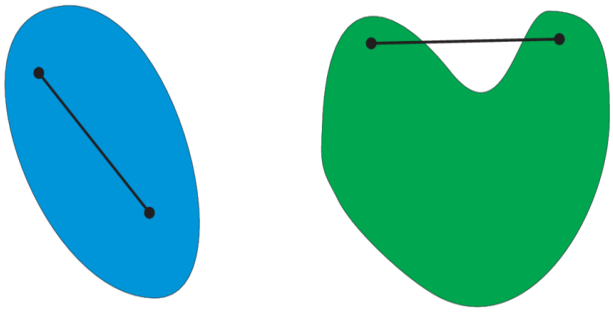

Fig. 1.1: Illustration of a convex (blue) and a non-convex (green) set.

Definition 1.1.2. Let $\sigma := \{x_i\}_{i \in I}$ be a subset of $\mathbb{R}^n$, where I is a finite set of indexes, we define σ to be *affinely independent* if $\{x_0 - x_i\}_{i \in I\text{-}\{0\}}$ is linearly independent.

We show now that the definition of affine independence of $\sigma = \{x_i\}_{i \in I} \subset \mathbb{R}^n$ is independent of the choice of x_0.

Proposition 1.1.3. *Let $\sigma := \{x_i\}_{i\in I}$ be a finite subset of $\mathbb{R}^n$, let $j \in I$ then, if $\{x_j - x_i\}_{i\in I\text{-}\{j\}}$ is linearly independent, also $\{x_0 - x_i\}_{i\in I\text{-}\{0\}}$ is.*

Proof. If $j = 0$ the statement is trivially true. Let $j \neq 0$ and $\lambda_i \in \mathbb{R}$ for all $i \neq j$, then

$$\sum_{i\in I\text{-}\{j\}} \lambda_i(x_j - x_i) = 0 \Rightarrow \lambda_i = 0 \quad \forall i \in I\text{-}\{j\}.$$

Let then $\mu_i \in \mathbb{R}$ for all $i \neq 0$, and suppose

$$\sum_{i\in I\text{-}\{0\}} \mu_i(x_0 - x_i) = (x_0 - x_j)\sum_{i\in I\text{-}\{0\}} \mu_i + \sum_{i\in I\text{-}\{0\}} \mu_i(x_j - x_i) = 0.$$

If we define $\mu_0 := -\sum_{i\in I\text{-}\{0\}} \mu_i$ we have that

$$0 = \sum_{i\in I} \mu_i(x_j - x_i) = \sum_{i\in I\text{-}\{j\}} \mu_i(x_j - x_i) \Rightarrow \mu_i = 0 \quad \forall i \in I\text{-}\{j\},$$

which proves our proposition. the definition of affine ind the definition of affine independence is well stated. □

Definition 1.1.4. Let $\sigma := \{x_i\}_{i\in I}$ be a finite subset of $\mathbb{R}^n$, we define the *convex set generated by σ* to be the smallest convex set containing X according to the inclusion relation. We shall denote this set by $[\sigma]$ and call it *convex hull* of σ.

Since the intersection of convex sets is convex, the convex set generated by σ can be equivalently defined as the intersection of all convex sets containing σ.

Theorem 1.1.5. *Let $\sigma := \{x_i\}_{i\in I}$ be a finite subset of $\mathbb{R}^n$, if σ is affinely independent then the convex set generated by σ is*

$$[\sigma] = \{\sum_{i\in I} \lambda_i x_i : \lambda_i \geq 0, \sum_{i\in I} \lambda_i = 1\}.$$

Furthermore for any point $x \in [\sigma]$ we have that

$$x = \sum_{i\in I} \lambda_i x_i = \sum_{i\in I} \mu_i x_i \Rightarrow \lambda_i = \mu_i \forall i \in I,$$

where $\lambda_i, \mu_i \geq 0$ and $\sum_{i\in I} \lambda_i = \sum_{i\in I} \mu_i = 1$.

Proof. Let $C := \{\bigcap_\alpha C_\alpha : \sigma \subset C_\alpha, C_\alpha \text{ convex}\}$, we divide the proof in three steps:

(i) $C \subset [\sigma]$.
This is true if $[\sigma]$ is convex and contains σ. The proof that it contains σ is trivial. In fact for every vertex $x_j = \sum_{i\in I} \delta_{ij} x_i$, and $\sum_{i\in I} \delta_{ij} = 1$.
To prove that it is convex we chose two points $a = \sum_{i\in I} a_i x_i, b = \sum_{i\in I} b_i x_i$
where $a_i, b_i \geq 0 \; \forall i \in I$ and $\sum_{i\in I} a_i = \sum_{i\in I} b_i = 1$. For $t \in [0,1]$

$$ta + (1-t)b = t\sum_{i\in I} a_i x_i + (1-t)\sum_{i\in I} b_i x_i = \sum_{i\in I} (ta_i + (1-t)b_i)x_i.$$

Since $ta_i + (1-t)b_i \geq 0$ and $\sum_{i\in I}(ta_i + (1-t)b_i) = t\sum_{i\in I} a_i + (1-t)\sum_{i\in I} b_i = 1$ for all $i \in I$, our statement is proven.

(ii) $[\sigma] \subset C$.
If all but one the λ_i are zero certaintly $\sum_{i\in I} \lambda_i x_i \in C$, since C contains all the vertexes. The inuctive hypothesis, by relabeling, is that if the first $\lambda_0, ..., \lambda_{n-1}$ are non-zero, hence

not even 1, then $\sum_{i\in I}\lambda_i x_i \in C$. We want to show that whenever $\lambda_0,\dots,\lambda_n$ are non-zero then also $\sum_{i\in I}\lambda_i x_i \in C$, since $\lambda_n \neq 1$ we have that

$$\sum_{i\in I}\lambda_i x_i = \sum_{i=0}^{n}\lambda_i x_i = \lambda_n x_n + \sum_{i=0}^{n-1}\lambda_i x_i = \lambda_n x_n + (1-\lambda_n)\sum_{i=0}^{n-1}\frac{\lambda_i}{1-\lambda_n}x_i.$$

Since $\sum_{i=0}^{n-1}\frac{\lambda_i}{1-\lambda_n} = 1$, for the inductive hypothesis $\sum_{i=0}^{n-1}\frac{\lambda_i}{1-\lambda_n}x_i \in C$. Also the vertex x_n is contained in C by definition, therefore, being C convex and $\lambda_n \in [0,1]$, it follows that

$$\lambda_n x_n + (1-\lambda_n)\sum_{i=0}^{n-1}\frac{\lambda_i}{1-\lambda_n}x_i \in C.$$

Accordingly $\sum_{i\in I}\lambda_i x_i \in C$, by induction we conclude the proof.

(iii) $\sum_{i\in I}\lambda_i x_i = \sum_{i\in I}\mu_i x_i \Rightarrow \lambda_i = \mu_i \forall i \in I$.
Let $\sum_{i\in I}\lambda_i x_i = \sum_{i\in I}\mu_i x_i$, then also $x_0\sum_{i\in I}\lambda_i + \sum_{i\in I}\lambda_i(x_i - x_0) = x_0\sum_{i\in I}\mu_i + \sum_{i\in I}\mu_i(x_i - x_0)$, and since both λ_i and μ_i are normalised we have that

$$\sum_{i\in I}(\lambda_i - \mu_i)(x_0 - x_i) = \sum_{i\in I\text{-}\{0\}}(\lambda_i - \mu_i)(x_0 - x_i) = 0 \Rightarrow \lambda_i = \mu_i \quad \forall i \in I\text{-}\{0\},$$

because of the affine independence. □

Definition 1.1.6. We define a *p-simplex* $[\sigma]$ to be the convex hull of an affinely independent set $\sigma := \{x_i\}_{i\in I} \subset \mathbb{R}^n$, where $p = |I| - 1$ is called dimension of the p-simplex.

Theorem 1.1.5 gives us the possibility to represent a point in a simplex $[\sigma]$ via a finite set of real parameters defined in the range $[0,1]$ and satisfying the normalisation condition $\sum_{i\in I}\lambda_i = 1$. Such parameters are called *baricentric coordinates* of $[\sigma]$.

The points in σ are called *vertexes* of the simplex $[\sigma]$, accordingly we define the vertex set of a simplex $[\sigma]$ to be $Vert([\sigma]) = \sigma$.

Definition 1.1.7. Let $[\sigma]$ be a p-simplex and $p, t \in \mathbb{N}$, we say that another t-simplex $[\tau]$ is a *face* of $[\sigma]$ or equivalently that $[\sigma]$ is a *coface* of $[\tau]$, and we write $[\tau] \leq [\sigma]$, if $\tau \subset \sigma$, where $t \leq p$.

Now we are ready for our main definitions.

Definition 1.1.8. We define a *simplicial complex* $\mathcal{G}$ to be a collection of simplexes such that

(i) if any simplex $[\tau] \leq [\sigma]$ and $[\sigma] \in \mathcal{G}$, then $[\tau] \in \mathcal{G}$,

(ii) if $[\sigma], [\tau] \in \mathcal{G}$, then $[\sigma] \cap [\tau] \in \mathcal{G}$.

Figure 1.2 represents a simplicial complex, while Figure 1.3 represents a collection of simplexes which is not a simplicial complex.

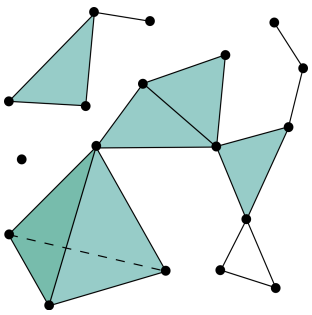

Fig. 1.2: Example of simplicial complex.

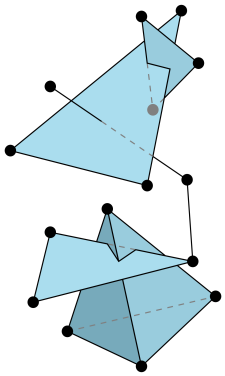

Fig. 1.3: Set of simplexes which is not a simplicial complex.

In fact, we can see in Figure 1.3 that the intersection property of simplicial complexes is not satisfied.

1.2 Abstract Simplicial Complexes

In section 1.1 we studied simplicial complexes as subsets of $\mathbb{R}^n$. From now we shall call them *geometric simplicial complexes*. Although this approach provides simplicial complexes with the topology inherited from the metric space, it hides the power of simplicial complexes to describe networks and interactions which exist independently of that topology. To make this distinction clear, we will treat simplicial complexes as a realization of more abstract objects called *abstract simplicial complexes*. A richer discussion of abstract simplicial complexes can be found in [4] at 3.1 or in [8] at 7.3.

Definition 1.2.1. Let $\mathcal{V}$ be a finite set, we define an *abstract simplicial complex* $\mathscr{A}$ to be a family of non empy subsets of $\mathcal{V}$ such that:

(i) if $v \in \mathcal{V}$, then $\{v\} \in \mathscr{A}$,

(ii) if $\sigma \in \mathscr{A}$ and $\tau \subset \sigma$, then $\tau \in \mathscr{A}$.

We call the members of this family *abstract simplexes*.

We call $\mathcal{V}$ the *vertex set* of $\mathscr{A}$ and denote it by $Vert(\mathscr{A})$; since the vertex set is finite, every abstract simplex is finite, therefore we can use the notation $\sigma = \{v_i\}_{i\in I}$, to denote a simplex in $\mathscr{A}$.

Definition 1.2.2. Let $\mathscr{A}$ be an abstract simplicial complex and $\mathscr{G}$ a geometric simplicial complex, if for all $\{x_i\}_{i\in I} \in \mathscr{A}$ also $[x_i]_{i\in I} \in \mathscr{G}$ we say that $\mathscr{G}$ is a *geometric realization* of $\mathscr{A}$.

While every geometric simplicial complex can be thought as a geometric realization of an abstract simplicial complex, the existence of a geometric realization for an arbitrary abstract simplicial complex is not trivial at all.

Theorem 1.2.3. *Let $\mathscr{A}$ be an n-dimensional abstract simplicial complex, then it admits a geometric realization in $\mathbb{R}^{2n+1}$.*

A proof of this theorem can be found in [4] at 3.1.

Both for abstract and geometric simplicial complexes one can define maps called *simplicial maps*. We obtain a category whose equivalences are called isomorphisms. A short discussion

of category theory can be found in Appendix A.

In the following sections we shall use abstract simplicial complexes, which can be always thought geometrically in the appropriate $\mathbb{R}^{2d+1}$.

1.3 Simplicial Homology

Homology and cohomology are key concepts in algebraic topology. We shall discuss homology theory to the extent that allows us to define the laplacian operator on simplicial complexes, for supplementary readings see [9] at 6.1 or [5] at 2.1. First we want to equip our simplicial complexes with an orientation. So far we have considered the simplex $\{x_i\}_{i\in I}$ up to reorderings of the index set I, but in most applications this is not the case.

We denote by $\{\sigma_i\}_{i\in I}$, where $I = \{1,...,n_p\}$, the set of all p-simplexes of the complex each one with a choice of orientation.

Definition 1.3.1. Let $\mathscr{A}$ be a simplicial complex, we define the *group of p-chains*

$$\mathscr{C}_p(\mathscr{A},\mathbb{R}) := \{\sum_{i\in I} \lambda_i\{\sigma_i\} : \lambda_i \in \mathbb{R}\}.$$

In the definition 1.3.1 the sum $\sum_{i\in I}\lambda_i\{\sigma_i\}$ is the formal sum of the free abelian group on $\mathbb{R}$, see [6] at 1.7. This can be generalized from $\mathbb{R}$ to an arbitrary abelian group, nevertheless, for the most of the applications, the groups $\mathbb{Z},\mathbb{R},\mathbb{Z}_2$ are considered.

Definition 1.3.2. Let $\mathscr{A}$ be a simplicial complex, we define the *group of oriented p-chains* as the group $\mathscr{C}_p(\mathscr{A},\mathbb{R})$ subject to $\{x_i\}_{i\in I} = sgn(\pi)\{x_i\}_{i\in\pi(I)}$, where $\pi : I \to I$ is any permutation of the index set. We shall denote this group as $C_p(\mathscr{A},\mathbb{R})$, and its chains with the notation $|\sigma\rangle$.

To keep our notation light we shall write C_p instead of $C_p(\mathscr{A},\mathbb{R})$. The constraints due to the permutations in definition 1.3.2 simply implies that whenever we swap two vertexes in the simplex we must multiply by (-1).

Proposition 1.3.3. *C_p is an abelian group.*

A particularly relevant role in homology theory is played by the *boundary map*. First we define the boundary of an oriented simplex.

Definition 1.3.4. Let $|\sigma\rangle = |x_0,...,x_{p+1}\rangle$ be an oriented $(p+1)$-simplex. The boundary $\partial|\sigma\rangle$ of $|\sigma\rangle$ is the p-chain defined by

$$\partial_{p+1}|\sigma\rangle := \sum_{i=0}^{p+1}(-1)^i|x_0,...,\widehat{x_i},...,x_{p+1}\rangle$$

where the $\widehat{\ }$ over a symbol means that symbol is deleted.

Remark. Note that whenever we are able to construct a geometric realization for the oriented simplicial complex, the set $\bigcup_{i=0}^{p+1}[x_0,...,\widehat{x_i},...,x_{p+1}]$ is the topological boundary of $[\sigma]$.

Furthermore we are able to extend the boundary from simplexes to chains.

Definition 1.3.5. We define the *boundary map* $\partial_{p+1} : C_{p+1} \to C_p$ to be the group homomorphism defined by

$$\partial_{p+1}(\sum_{i\in I}\lambda_i|\sigma_i\rangle) := \sum_{i\in I}\lambda_i\partial_{p+1}|\sigma_i\rangle.$$

An important property of boundary maps is the following, which we shall denote as *homology lemma*.

Lemma 1.3.6. *The boundary maps satisfy* $\partial_p \circ \partial_{p+1} = 0$.

Proof. Since the boundary maps are linear it is sufficient to check this on the generators. Let $\partial_{p+1}|x_0,\dots,x_{p+1}\rangle = \sum_{i=0}^{p+1}(-1)^i|x_0,\dots,\widehat{x_i},\dots,x_{p+1}\rangle$ then

$$(\partial_p \circ \partial_{p+1})|x_0,\dots,x_{p+1}\rangle = \sum_{j=0,j\neq i}^{p+1}\sum_{i=0}^{p+1}(-1)^{i+j}|x_0,\dots,\widehat{x_i},\dots,\widehat{x_j},\dots,x_{p+1}\rangle = 0. \qquad \square$$

The homology lemma is necessary to define the *homology group*. The homology group is , intuitively, the space of cycles that are not boundaries. In fact, without the homology lemma the quotient would not well defined since $im\partial_{p+1} \subset ker\partial_p$ would not be satisfied.

Definition 1.3.7. We define the *p-homology group* to be

$$H_p := \frac{\ker \partial_p}{\operatorname{im} \partial_{p+1}},$$

where $\operatorname{im}\partial_{p+1}$ is called the group of simplicial *p-cycles* and $\ker\partial_p$ is called the group of simplicial *p-boundaries*.

1.4 Simplicial Cohomology

In order to define data on simplicial complexes we are interested in studying the dual of the chains $|\sigma\rangle$, which we shall call cochains $\langle\sigma|$.

Definition 1.4.1. Let C_p be the group p-chains, we define the group of *p-cochains* to be

$$C^p := Hom(C_p, \mathbb{R}).$$

Proposition 1.4.2. *C^p is an abelian group.*

Proposition 1.4.3. *The homomorphims $\{\langle\sigma_i| : C_p \to \mathbb{R}\}$ such that $\langle\sigma_i|\sigma_j\rangle =: \langle\sigma_i|(|\sigma_j\rangle) = \delta_{ij}$, form a basis of C^p.*

As well as with for chains, we have also for cochains a sequence of homeomorphisms called *coboundary maps*. The coboundary maps are defined to be the dual of the boundary maps, hence satisfying a dual version of the homology lemma called *cohomology lemma*.

Definition 1.4.4. The dual of the boundary maps which we shall call *coboundary maps*, is the group homomorphsm defined by

$$d_{p+1} : C^p \to C^{p+1} \quad d_{p+1}\langle\sigma| := \langle\sigma|\partial_{p+1} \quad \forall \langle\sigma| \in C^p.$$

Therefore $d_{p+1}\langle\sigma|(|\tau\rangle) = \langle\sigma|(\partial_{p+1}|\tau\rangle) \quad \forall |\tau\rangle \in C_p$.

The proof of $d_{p+1} \circ d_p = 0$ follow directly from the homology lemma. The cohomology lemma allows us to define the *cohomology group*.

Definition 1.4.5. We define the *p-cohomology group* to be

$$H^p := \frac{\ker d_{p+1}}{\operatorname{im} d_p},$$

where $\operatorname{im} d_{p-1}$ is the group of simplicial p-cocycles and $\ker d_p$ is the group of simplicial p-coboundaries.

Chapter 2

Graphs

An important branch of geometric deep learning is graph representation learning. In this section we will therefore introduce graphs and graphs laplacians to build a mathematical background for graph neural networks.

2.1 Definition of Graph

Graphs are used to encode a wide variety of data, from social networks friendships to maps. We can actually consider graphs as the first topological notion introduced in mathematics. Their introduction dates back to the XVIII century. Euler noticed that in Koningsberg, there was no path that allowed to cross all seven bridges just once. Hence, he started to simplify the problem to approach it in a more mathematical way.

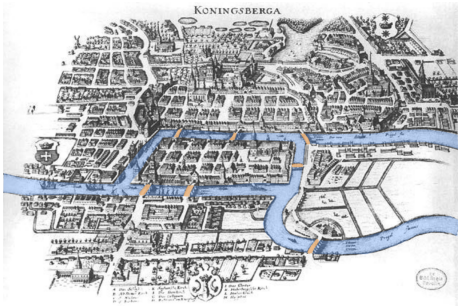

Fig. 2.1: The city of Konigsberg and the seven bridges.

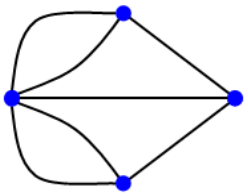

Fig. 2.2: Graph representing Konigsberg's seven bridges.

This led Euler to formulate an abstract approach to the problem that led to its solution, presented to the St. Petersburg Academy in 1735.
In chapter 1 we presented abstract simplicial complexes, in this chapter we shall discuss in detail 1-dimensional abstract simplicial complexes and how they are related to graphs. More information on graphs can be found in [1].

Definition 2.1.1. An *undirected graph* $\mathcal{G}$ is an orderd pair $(V(\mathcal{G}),E(\mathcal{G}))$, consisting of a set $V(\mathcal{G})$ of *vertices* and a set $E(\mathcal{G})$, disjoint from $V(\mathcal{G})$, of *edges*, together with an incidence function $\psi_{\mathcal{G}} : E(\mathcal{G}) \to V(\mathcal{G}) \times V(\mathcal{G})$ that associates to each edge of $\mathcal{G}$ an unordered pair of (not necessarily distinct) vertices of $\mathcal{G}$. If e is an edge and u and v are vertices such that $\psi_{\mathcal{G}}(e) = \{u,v\}$, then e is said to *join* u and v, and the vertices u and v are called the *ends* of e. A graph where the ordering of the ends of an edge is meaningful is said to be a *directed* graph.

Fig. 2.3 represents an undirected graph with nine vertices and seven edges.

Definition 2.1.2. Let $\mathscr{G}$ be an undirected graph, we define the *adjacency matrix* A of $\mathscr{G}$ as follows, A is a $n \times n$ symmetric matrix, where n is the number of vertices, such that

$$a_{ij} = \begin{cases} 1 & \text{if there exists an edge joining } i \text{ with } j \\ 0 & \text{otherwise} \end{cases}.$$

Remark. We can also define the adjacency matrix for directed graphs. For a directed graph the adjacency matrix is not symmetric.

Definition 2.1.3. A graph is *simple* if it has neither different edges connecting the same pair of verices nor loops, i.e. there is a unique edge joining two different vertices and no vertex is joined to itself by an edge. This implies $a_{ii} = 0$ in the adjacency matrix.

Much of graph theory is concerned with the study of simple graphs.

Observation. We can view a *simple undirected graph* $\mathscr{G}$ as 1-dimensional abstract simplicial complex. In fact we can construct a family of sets $\mathscr{A}$ as in definition 1.2.1 in the following way.

$$\mathscr{A} = V(\mathscr{G}) \cup \{\{i,j\} \in V(\mathscr{G}) \times V(\mathscr{G}) : \exists e \in E(\mathscr{G}) \quad \psi_{\mathscr{G}}(e) = \{i,j\}\}.$$

It is easy to see that this is an abstract simplicial complex of dimension 1, one can readily verify the properties in def. 1.2.1.

Viewing $\mathscr{G}$ as a simplicial complex we recall that C_0 and C_1 are the groups of 0 and 1-chains respectively. We denote by $|i\rangle$ the vertex i as a group element of C_0, and we denote $|i,j\rangle$ the edge joining $|i\rangle$ to $|j\rangle$. Since C_1 is a group, whenever $|i,j\rangle \in C_1$ then also $|j,i\rangle = -|i,j\rangle \in C_1$.
A canonical basis for C_0 is the set of oriented 0-chains consisting on only one oriented 0-simplex, similarly for C_1.

2.2 Homology and Cohomology on Graphs

SInce a graph is a 1-diemnsional simplicial complex we can define only one boundary ∂_1 and one coboundary $\partial_1^\dagger$. We briefly recall the definitions in this context.

Definition 2.2.1. Let $\mathscr{G}$ be a graph, we define the *boundary* $\partial_1 : C_1 \to C_0$ by

$$\partial_1 |i,j\rangle := |i\rangle - |j\rangle \quad \forall |i,j\rangle \in C_1.$$

We define the *coboundary* $\partial_1^\dagger : C_0 \to C_1$ by

$$\partial_1^\dagger |i\rangle := \sum_{j : |i,j\rangle \in C_1} |i,j\rangle.$$

We define the *divergence* $d_1 : C^0 \to C^1$ by

$$d_1 \langle i| = \langle i|\partial_1,$$

where the cochain $\langle i|\partial_1$ is the dual of the chain $\partial_1^\dagger |i\rangle$, therefore we can represent d_1 on cochain via $\partial_1^\dagger$ on chains.

We define $d_1 = (\partial_1)^*$ and $d_1^\dagger = (\partial_1^\dagger)^*$ and we identify $C_p = C^p$ with the above choice of canonical basis.

2.3 Laplacian Operators on Graphs

On a graph that admits non trivial gradient and divergence, we can define a non trivial laplacian operator on cochains as the gradient of the divergence. Since we are going to represent it on chains we have that

$$d_1 d_1^\dagger \langle i| = \langle i|\partial_1^\dagger \partial_1,$$

which is the adjoint of $\partial_1 \partial_1^\dagger |i\rangle$.

Definition 2.3.1. Let $\mathscr{G}$ be a graph, we define the 0-*laplacian* $\Delta_0 : C_0 \to C_0$ by

$$\Delta_0 := \partial_1 \partial_1^\dagger.$$

One interesting propertiy of the 1-laplacian is that the dimension of its kernel equals the connected components of the graph. To illustrate some of the properties of graphs we will refer to the graph in Figure 2.3.

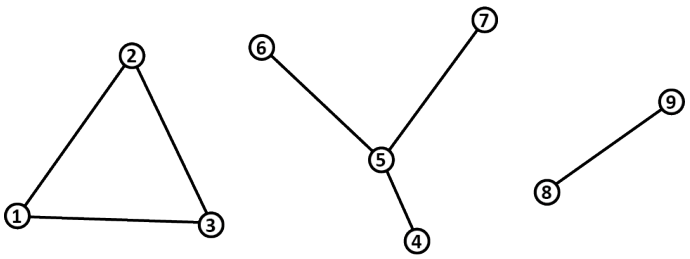

Fig. 2.3: The graph.

$$\begin{pmatrix}
0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}.$$

Fig. 2.4: Adjacency matrix of the graph.

Example 2.3.2. Let $\mathscr{G}$ be the graph in Figure 2.3, then the 0-laplacian expressed in terms of the canoncal vertex basis $\{|i\rangle\}_{i\in I}$ is

$$\begin{pmatrix}
2 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & -1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 3 & -1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1
\end{pmatrix}.$$

The laplacian matrix is calculated as follows

$$\Delta_0|1\rangle = \partial_1|1,2\rangle + \partial_1|1,3\rangle = 2|1\rangle - |2\rangle - |3\rangle,$$

$$\Delta_0|2\rangle = \partial_1|2,1\rangle + \partial_1|2,3\rangle = -|1\rangle + 2|2\rangle - |3\rangle,$$

$$\Delta_0|3\rangle = \partial_1|3,1\rangle + \partial_1|3,2\rangle = -|1\rangle + -|2\rangle + 2|3\rangle,$$

$$\Delta_0|4\rangle = \partial_1|4,5\rangle = |4\rangle - |5\rangle,$$

$$\Delta_0|5\rangle = \partial_1|5,4\rangle + \partial_1|5,6\rangle + \partial_1|5,7\rangle = -|4\rangle + 3|5\rangle - |6\rangle - |7\rangle,$$

$$\Delta_0|6\rangle = \partial_1|6,5\rangle = -|5\rangle + |6\rangle,$$
$$\Delta_0|7\rangle = \partial_1|7,5\rangle = -|5\rangle + |7\rangle,$$
$$\Delta_0|8\rangle = \partial_1|8,9\rangle = |8\rangle - |9\rangle,$$
$$\Delta_0|9\rangle = \partial_1|9,8\rangle = -|8\rangle + |9\rangle.$$

Three invariant subspaces emerge from the laplacian, that determine three different laplacians, namely

$$\Delta_0 = \Delta_0^{\mathscr{A}} \oplus \Delta_0^{\mathscr{B}} \oplus \Delta_0^{\mathscr{C}}.$$

Furthermore any of those three blocks has a 1-dimensional kernel, in fact the dimensional equiations for the laplacians are

$$dim(ker\Delta_0^{\mathscr{A}}) = dimC_0(\mathscr{A}) - rank\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} = 3 - rank\begin{pmatrix} 2 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} = 3 - 2 = 1,$$

$$dim(ker\Delta_0^{\mathscr{B}}) = dimC_0(\mathscr{B}) - rank\begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} = 4 - rank\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 2 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} = 4 - 3 = 1,$$

$$dim(ker\Delta_0^{\mathscr{C}}) = dimC_0(\mathscr{C}) - rank\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = 2 - rank\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} = 2 - 1 = 1.$$

We can also define an higher dimensional laplacian on the graph.

Definition 2.3.3. Let $\mathscr{G}$ be a graph, we define the 1-*laplacian* $\Delta_1 : C_1 \to C_1$ by

$$\Delta_1 := \partial_1^\dagger \partial_1.$$

One intersting propertiy of the 1-laplacian is that the dimension of its kernel equals the number of independent cycles.

Example 2.3.4. In fact we can expand $\Delta_1^{\mathscr{A}} := \partial_1^\dagger \partial_1$ the basis $\{|1,2\rangle, |2,3\rangle, |3,1\rangle\}$ as

$$\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}.$$

The laplacian matrix is calculated as follows

$$\Delta_1^{\mathscr{A}}|1,2\rangle = \partial_1^\dagger(|1\rangle - |2\rangle) = 2|1,2\rangle - |3,1\rangle - |2,3\rangle,$$
$$\Delta_1^{\mathscr{A}}|2,3\rangle = \partial_1^\dagger(|2\rangle - |3\rangle) = -|1,2\rangle + 2|2,3\rangle - |3,1\rangle,$$
$$\Delta_1^{\mathscr{A}}|3,1\rangle = \partial_1^\dagger(|3\rangle - |1\rangle) = -|1,2\rangle + 2|3,1\rangle - |2,3\rangle.$$

We can notice that Δ_2 has a 1-dimensional kernel.

$$dim(ker\Delta_1^{\mathscr{A}}) = dimC_1(\mathscr{A}) - rank\begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} = 3 - rank\begin{pmatrix} 2 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} = 3 - 2 = 1.$$

Since the sum of the three raws is 0 we can say that the only linearly independent 1-cycle is $|1,2\rangle + |2,3\rangle + |3,1\rangle$.

2.4 Heat Equation on Graphs

An relevant differential equation in physics and mathematics is the so called *heat equation*

$$\frac{df}{dt} = -\Delta f.$$

Proposition 2.4.1. *Let $|\rho\rangle \in C_0$ be a time dependent chain on our graph representing a density, and let the mass $M = \sum_{|i\rangle \in C_0} \langle i|\rho\rangle$ be a conserved quantity, then given a flux for $|\rho\rangle$, namely a time dependent $|j\rangle \in C_1$ such that $|j\rangle = \partial_1^\dagger |\rho\rangle$, then $|\rho\rangle$ satisfies the discrete heat equation*

$$\frac{\partial}{\partial t}|\rho\rangle = -\Delta_0|\rho\rangle.$$

Proof. Since the mass M is a conserved quantity, we need the flux on the coboundary of any vertex set A added to the variation of the mass within A to be zero, hence a local continuity equation. The vertex set A can be seen as the chain $|A\rangle = \sum_{i\in A} |i\rangle$, with this notation we have that the previous sentece reads as

$$\frac{d}{dt}\langle A|\rho\rangle + \langle A|\partial_1|j\rangle = 0.$$

In fact, $\langle A|\partial_1|j\rangle = \langle j|\partial_1^\dagger|A\rangle$ is the flux evaluated on the coboundary of A $\partial_1^\dagger|A\rangle$.
It follows that

$$\frac{d}{dt}\langle A|\rho\rangle + \langle A|\partial_1|j\rangle = \sum_{i\in A}\langle i|(\frac{d}{dt}|\rho\rangle + \partial_1|j\rangle) = 0,$$

since $\langle i|$ is not time dependent, which leads to the local continuity equation

$$\frac{d}{dt}|\rho\rangle + \partial_1|j\rangle = 0.$$

Inserting the definition of flux in the continuity equation we obtain the discrete heat equation on graphs

$$\frac{d}{dt}|\rho\rangle + \partial_1\partial_1^\dagger|\rho\rangle = \frac{d}{dt}|\rho\rangle + \Delta_0|\rho\rangle = 0.$$

□

We can therefore interpret the heat equation on graphs as a counter-gradient diffusion of some conserved quantity. Let us see what we can say about the solution of this equation on graphs.

Proposition 2.4.2. *Let $\frac{d}{dt}|\psi\rangle = -\Delta_0|\psi\rangle$ and $|\psi\rangle|_{t=0} = |\psi_0\rangle$, then*

$$|\psi\rangle = \sum_i |e_i\rangle e^{-\lambda_i t}\langle e_i|\psi_0\rangle,$$

where $|e_i\rangle$ and λ_i are the orthonornmal eigenchains and the eigenvalues of the 0-laplacian respectively, i.e. $\Delta_0|e_i\rangle = |e_i\rangle\lambda_i$.

Proof. Since Δ_0 is selfadjoint and therefore admits the expansion $\Delta_0 = \sum_i |e_i\rangle\lambda_i\langle e_i|$, where $|e_i\rangle$ and λ_i are the orthonornmal eigenchains and the eigenvalues of the 0-laplacian respectively, i.e. $\Delta_0|e_i\rangle = |e_i\rangle\lambda_i$. The discrete heat equation is linear and with constant coefficients and therefore it admits a solution of the type $|\psi\rangle = e^{-\Delta_0 t}|\psi_0\rangle$. Since $[\Delta_0, e^{-\Delta_0 t}] = 0$, these two operators share the eigenchains and therefore we also have that

$$e^{-\Delta_0 t} = \sum_i |e_i\rangle e^{-\lambda_i t}\langle e_i|,$$

hence the solution. □

From this solution we can notice that for $t \to \infty$, the only terms left are those with eigenvalue 0, namely the connected components. However, because of the term $\langle e_i|\psi_0\rangle$, only the connected components on which $|\psi_0\rangle$ is not zero everywhere remain.

Example 2.4.3. Let's for instance take the graph in Fig. 2.3. On this graph we want to solve the heat equation given the initial condition $|\psi_0\rangle = |2\rangle + |8\rangle$. Since the subgroup generated by $\{|4\rangle, |5\rangle, |6\rangle, |7\rangle\}$ has a null projection on the initial condition, being an invariant subgroup under Δ_0 and $e^{-\Delta_0 t}$, we have that those vertices do not appear in the final solution. In order to write the solution we need to diagonalize the other two blocks of the 0-laplacian. For the block of $\{|1\rangle, |2\rangle, |3\rangle\}$, we have

$$\Delta_0 \frac{|1\rangle + |2\rangle + |3\rangle}{\sqrt{3}} = 0,\ \Delta_0 \frac{|3\rangle - |1\rangle}{\sqrt{2}} = 3\frac{|3\rangle - |1\rangle}{\sqrt{2}},\ \Delta_0 \frac{|2\rangle - |1\rangle}{\sqrt{2}} = 3\frac{|2\rangle - |1\rangle}{\sqrt{2}}.$$

For the block of $\{|8\rangle, |9\rangle\}$, we have

$$\Delta_0 \frac{|8\rangle + |9\rangle}{\sqrt{2}} = 0,\ \Delta_0 \frac{|8\rangle - |9\rangle}{\sqrt{2}} = 2\frac{|8\rangle - |9\rangle}{\sqrt{2}}.$$

Hence the solution takes the form

$$|\psi\rangle = \frac{|1\rangle + |2\rangle + |3\rangle}{\sqrt{3}} \frac{\langle 1| + \langle 2| + \langle 3|}{\sqrt{3}}(|2\rangle + |8\rangle) + \frac{|2\rangle - |1\rangle}{\sqrt{2}} e^{-3t} \frac{\langle 2| - \langle 1|}{\sqrt{2}}(|2\rangle + |8\rangle) +$$
$$\frac{|8\rangle + |9\rangle}{\sqrt{2}} \frac{\langle 8| + \langle 9|}{\sqrt{2}}(|2\rangle + |8\rangle) + \frac{|8\rangle - |9\rangle}{\sqrt{2}} e^{-2t} \frac{\langle 8| - \langle 9|}{\sqrt{2}}(|2\rangle + |8\rangle) =$$
$$= \frac{|1\rangle + |2\rangle + |3\rangle}{3} + \frac{|2\rangle - |1\rangle}{2} e^{-3t} + \frac{|8\rangle + |9\rangle}{2} + \frac{|8\rangle - |9\rangle}{2} e^{-2t}.$$

In the limit

$$\lim_{t\to\infty} |\psi\rangle = \frac{|1\rangle + |2\rangle + |3\rangle}{3} + \frac{|8\rangle + |9\rangle}{2},$$

we see that the conserved quantity that initally was 1 on both those connected components is in this limit equally shared among the vertices.

Chapter 3

Geometric Deep Learning

3.1 Group Equivariant Neural Networks

In many classification task the neural networks seem to learn some symmetries under which the classification is invariant. In some simple situations, where these symmetries are obvious and understandable by humans, we hope to find a neural architecture that already knows the symmetry of the data before starting to learn. An approach was proposed in [2] with group equivariant convolutional networks. Before discussing those particular networks we analyse the meaning of group equivariance.

3.2 Convolutional Neural Networks

In section 3.1 we discussed group equivariant convolutional networks, now we analyse standard convolutional networks as shift equivariant networks.

3.3 Simplicial Neural Networks

In chapter 1 we discussed laplacian operators, in this section we use the laplacian eigenchains to define convolutional layers as in [3] in a so called *simplicial neural network*.
Let $\{|i\rangle\}_{i\in I}$ be the canonical basis of chains for C_p, namely the chains corresponding to the p-simplexes of the simplicial complex. Nevertheless can also choose the laplacian eigenchains $\{|e_i\rangle\}_{i\in I}$, such that $\Delta_p|e_i\rangle = \lambda_i|e_i\rangle$, as a basis for C_p. This change of basis is reflected on a change of coordinates called *simplicial Fourier transform*.

Definition 3.3.1. Let $|f\rangle \in C_p$ and $dimC_p = n_p$, we define the *simplicial Fourier transform* $\mathscr{F}_p : \mathbb{R}^{n_p} \to \mathbb{R}^{n_p}$ to be

$$(\langle i|f\rangle)_{i\in I} \mapsto (\langle e_i|f\rangle)_{i\in I},$$

where $I = \{1,...,n_p\}$.

This transform only defines a change of basis since $|f\rangle = \sum_{i\in I}\langle i|f\rangle|i\rangle = \sum_{i\in I}\langle e_i|f\rangle|e_i\rangle$, and therefore is invertible. We can in fact represent the simplicial Fourier transform with the matrix $F^{-1}_{ij} = F^{\dagger}_{ij} := \langle e_i|j\rangle$, and its inverse $F_{ij} := \langle i|e_j\rangle$. To define a convolution between two p-chains we use the famous convolution theorem $\mathscr{F}(f * \psi) = \mathscr{F}(f)\mathscr{F}(\psi)$.

Definition 3.3.2. Let $\{|e_i\rangle\}_{i\in I}$ be a basis such that $\Delta_p|e_i\rangle = \lambda_i|e_i\rangle$, let $|f\rangle, |\psi\rangle \in C_p$, we define the representatives of $|f * \psi\rangle$ on the laplacian eigenchains to be

$$\langle e_i|f * \psi\rangle := \langle e_i|f\rangle\langle e_i|\psi\rangle \quad \forall i \in I.$$

Therefore $|f * \psi\rangle = \sum_{i\in I}\langle e_i|f\rangle\langle e_i|\psi\rangle|e_i\rangle$.

The filters used in [3] are low degree polynomials in the frequency domain, for instance a filter defined by $\langle i|\psi_\mu\rangle = \sum_{n=0}^{N} \mu_n \lambda_i^n$. This way one can easily define the convolution on the canonical basis as

$$\langle i|f * \psi_\mu\rangle = \sum_{j\in I} \langle i|e_j\rangle\langle e_j|f * \psi_\mu\rangle = \sum_{n=0}^{N} \mu_n \langle i|(\sum_{j\in I} |e_j\rangle \lambda_j^n \langle e_j|)|f\rangle = \sum_{n=0}^{N} \mu_n \langle i|\Delta_p^n|f\rangle,$$

therefore $|f * \psi_\mu\rangle = \sum_{n=0}^{N} \mu_n \Delta_p^n |f\rangle$.

3.4 Message Passing Neural Networks

In the previous section we saw one possible definition of convolutional layers on graph and simplicial complexes [1], here we shall discuss on graphs a more general definition based on the diffusion equation on graphs.

[1] In general on any chain complex

Conclusion

Most of the deep learning techniques used today are based on models which learn a partition of the set of smooth functions defined on euclidean domains into human friendly equivalence classes. Although this approach has been successful in modern machine learning, it only deals with a really small set of domains. The goal of geometric deep learning is to extend this method to data defined on manifolds and simplicial complexes.
Convolution on euclidean domains is itself based on the translation invariance of such domains. In fact the convolution of a function $f : \mathbb{R}^n \to \mathbb{R}$, with some filter $g : \mathbb{R}^n \to \mathbb{R}$ is

$$(f * g)(x) = \langle f, g \circ T^{-1} \rangle_{L^2},$$

where $T : \mathbb{R}^n \to \mathbb{R}^n$ is a translation represented by the vector x. Moreover, one could also consider such a convolution to be defined on the translation group itself, represented by some $\mathbb{R}^n$. In the same way one can define a convolution on the group $(\mathbb{Z}, +)$ as $(a * b)_n = \sum_{i \in \mathbb{Z}} a_i b_{n-i}$. Similarly an intersting example is that of images, which are samples ad grids, any image can be thought as a function defined on the group $(\mathbb{Z} \times \mathbb{Z}, +)$. Such definitions of convolution operators are equivariant with respect to the action of the group they are defined upon. In image recognition translation equivariance is necessary, nevertheless the most common CNN's need to learn the rotations of the same filter as different filters in order to become rotation equivariant. Although manifolds and simplicial complexes are not in general groups, G-equivariant CNN's (see [2]) could be the key to reveal the secrets behind the success of such architectures.

Appendix A

Category Theory

Definition A.1. A *category* $\mathbf{C}$ consists of three ingerdients:

1. a class of *objects* $Obj(\mathbf{C})$,
2. sets of *morphisms* $Hom(A,B)$ for every ordered pair $(A,B) \in Obj(\mathbf{C}) \times Obj(\mathbf{C})$,
3. a composition $Hom(A,B) \times Hom(B,C) \to Hom(A,C)$, denoted by $(f,g) \mapsto f \circ g$ for every $A,B,C \in Obj(\mathbf{C})$, satisfying the following axioms:
   (i) the family of $Hom(A,B)$ is pairwise disjoint,
   (ii) the composition, when defined, is associative,
   (iii) for each $A \in Obj(\mathbf{C})$ there exists an *identity* $1_A \in Hom(A,A)$ such that for $f \in Hom(A,B)$ and $g \in Hom(C,A)$ we have that $1_A \circ f = f$ and $g \circ 1_A = g$.

Instead of writing $f \in Hom(A,B)$, we usually write $f : A \to B$.

Definition A.2. Let $\mathbf{A}$ and $\mathbf{C}$ be categories, a *functor* $T : \mathbf{A} \to \mathbf{C}$ is a function, that is,

(i) for each $A \in Obj(\mathbf{A})$ it assigns $TA \in Obj(\mathbf{C})$,

(ii) for each morphism $f : A \to A'$ it assigns a morphism $Tf : TA \to TA' \quad \forall A, A' \in Obj(\mathbf{A})$,

(iii) if f, g are morphisms in $\mathbf{A}$ for which $g \circ f$ is defined, then $T(g \circ f) = (Tg) \circ (Tf)$,

(iv) $T(1_A) = 1_{TA} \quad \forall A \in \mathbf{A}$.

The property (iii) of the previous definition actually defines what we shall call *covariant functors*. If instead we require $T(g \circ f) = (Tf) \circ (Tg)$, we are defining a *contravariant functor*.

Definition A.3. An *equivalence* in a category $\mathbf{C}$ is a morphism $f : A \to B$ for all $A, B \in Obj(\mathbf{C})$ for which there exists a morphism $g : B \to A$ such that $f \circ g = 1_B$ and $g \circ f = 1_A$.

Theorem A.4. *If $\mathbf{A}$ and $\mathbf{C}$ are categories and $T : \mathbf{A} \to \mathbf{C}$ is a functor of either variance, then whenever f is an equivalence on $\mathbf{A}$ then Tf is an equivalence on $\mathbf{C}$.*

Proof. We apply T to the equations $f \circ g = 1_B$ and $g \circ f = 1_A$, that for a covariant functor leads to $(Tf) \circ (Tg) = T(1_B) = 1_{TB}$ and $(Tg) \circ (Tf) = T(1_A) = 1_{TA}$. □

A category that will be used in the following section is the category of topological spaces and continuous functions.

Proposition A.5. *Topological spaces and continuous functions are a category **Top**, whose equivalences are called homeomorphisms.*

Other examples of categories can be found in [8] at 0.3 and in [7].

Appendix B

Dirac Notation

In the Dirac notation we use the kets $|\psi\rangle$ as vectors (group elements) and the bras $\langle\psi|$ as their dual. The bra-ket product $\langle\psi|\varphi\rangle = \langle\psi|(\varphi)$ is the bra evaluated on the ket, this is an inner product of $|\psi\rangle$ and $|\varphi\rangle$. In this notation the spectral theorem for selfadjoint operators reads as

$$A = \sum_i |a_i\rangle a_i \langle a_i|,$$

where $A|a_i\rangle = a_i|a_i\rangle$, $\langle a_i|a_j\rangle = \delta_i j$ and $\sum_i |a_i\rangle\langle a_i| = 1$. The relation $\sum_i |a_i\rangle\langle a_i| = 1$ is called completeness relation and is valid only if $\{|a_i\rangle\}$ is an orthonormal basis, 1 is the identity operator.

Bibliography

[1] J. A. Bondy and U. S. R. Murty. *Graph Theory with Applications*. Elsevier, 1976.

[2] Taco S. Cohen and Max Welling. *Group Equivariant Convolutional Networks*. 2016.

[3] S. Ebli, M. Defferrard, and G. Spreemann. *Simplicial Neural Networks*. 2020.

[4] Herbert Edelsbrunner and John Harer. *Computational Topology - an Introduction.* American Mathematical Society, 2010.

[5] Allen Hatcher. *Algebraic Topology*. Cambridge Univ. Press, 2000.

[6] Serge Lang. *Algebra*. Springer, New York, NY, 1955.

[7] Saunders MacLane. *Categories for the Working Mathematician*. 1971.

[8] Joseph J. Rotman. *An Introduction to Algebraic Topology*. Graduate Texts in Mathematics. Springer-Verlag New York, 1988.

[9] I. M. Singer and J. A. Thorpe. *Lecture Notes on Elementary Topology and Geometry*. Springer, New York, NY, 1976.